

Teresa Maria Martins Dos Santos Neves

**ANÁLISE DO DESEMPENHO DE CARTAS MULTIVARIADAS  
COM DADOS AUTOCORRELACIONADOS: CARTA  $T^2$  E  
CARTA DPCA**

Lisboa

2010

## Agradecimentos

Apesar de uma dissertação ser um trabalho individual, a que qualquer investigador está destinado, reúne sempre contributos de várias pessoas, pelo que expresso aqui os meus sinceros agradecimentos a todos os que contribuíram para este trabalho.

Deste modo gostaria em primeiro lugar agradecer a minha família, pais, irmãs, tios e primos o total, sincero e incondicional apoio para vencer mais uma etapa da minha formação. Deixo expresso as minhas desculpas pela impaciência, irritabilidade, pela ausência no papel familiar. Obrigada pelo vosso amor.

Ao Wagner Madaleno pelo carinho, amizade, paciência, solidariedade e apoio incondicional dado durante esta árdua tarefa.

A Professora Ana Sofia, orientadora da dissertação, agradeço o apoio, incentivo e interesse demonstrado pelo trabalho em desenvolvimento. Acima de tudo, obrigada pela amizade, pela disponibilidade sempre revelada, pelo incansável apoio moral durante esta jornada.

Às minhas amigas, agradeço-vos do fundo do meu coração toda a amizade, carinho, incentivo, pelo apoio moral, pela compreensão e pelas noites que passamos em frente ao computador. Adoro-vos.

## Resumo

As cartas de controlo multivariadas têm despertado, cada vez mais, a atenção das indústrias com processos modernos. Actualmente o grande desafio que se coloca a essas indústrias que tencionam recorrer a técnicas multivariadas prende-se com o facto de, em muitas situações, os dados a controlar exibirem autocorrelação (uma observação num dado instante depende de observações ocorridas em instantes anteriores).

Com esse intuito, o presente trabalho tem como objectivo principal o desenvolvimento de uma metodologia que permita estabelecer um estudo comparativo entre o desempenho de duas cartas multivariadas com dados autocorrelacionados, a carta  $T^2$  e a carta baseada na análise de componentes principais dinâmicas (DPCA), de forma a indicar qual apresenta o melhor desempenho, quando o processo é sujeito a perturbações numa ou mais variáveis em simultâneo. Este estudo é feito através de simulação e tem como medida de desempenho, o ARL (*Average Run Length*).

A metodologia proposta, na presente dissertação, foi aplicada a partir de modelos obtidos de dados reais e os resultados alcançados indicam que a carta DPCA apresenta o melhor desempenho, quando se pretende detectar alterações na média do processo. Deste modo, proporcionou-se, aos engenheiros do processo, uma óptima ferramenta de apoio à tomada de decisão, quando o processo é afectado por perturbações numa ou mais variáveis em simultâneo.

## Abstract

The multivariate control chart have been, more and more, catching the attention of the industries with modern processes. Nowadays the biggest challenge presented to those industries, meaning to evoke for the multifaceted techniques is due to the fact that in many situations the data to be controlled display autocorrelation (an observation in a certain instant depends on observations made in previous instants)

With that intention, the present essay main objective is the development of a methodology that permits the establishment of a comparative study of the performance between two multivariate chart with autocorrelated data,  $T^2$  - chart and chart based on the dynamic principal component analysis (DPCA), in a way that indicates which one presents the best performance, when the process undergoes perturbations in one or more variables simultaneously. This study is made through simulation and has as measurement of performance, the ARL (Average Run Length).

In the present dissertation, the proposed methodology was applied with model built with real data and the obtained results indicate that the DPCA chart presents the best outcome when the desire is to detect alterations in the means of the process. Therefore, it was offered to the engineers of the process an outstanding tool of support in the decision, when the process is affected with perturbations in one or more variables simultaneously.

## Lista de Abreviaturas e Símbolos

### Abreviaturas

ARIMA (p,d,q)	Modelo Autorregressivo de ordem p, Diferenciação de ordem d e de Médias Móveis de ordem q	<i>AutoRegressive Integrated Moving Average</i>
ARL	Número médio ao fim do qual se destaca uma situação fora de controlo	<i>Average Run Length</i>
ARX	Modelo autorregressivo com variável Exógena	<i>Autorregressive Model with Exogene Variable</i>
AR (p)	Modelo Autorregressivo de ordem p	<i>Autorregressive Model</i>
CP	Componente Principal	
Carta CCC	Carta de Causas Comuns	<i>Commum-Cause Chart</i>
Carta CUSUM	Carta de Somas Acumuladas	<i>Cumulative Sums Chart</i>
Carta EWMA	Carta de Média Móvel Exponencialmente Amortecida	<i>Exponential Weight Moving Range Chart</i>
Carta EWMAST	Carta de Média Móvel Exponencialmente Amortecida	
Carta MCEWMA	Carta EWMA com limites móveis	<i>Moving Center Line EWMA Chart</i>
Carta MEWMA	Carta EWMA Multivariada	<i>Multivariate EWMA Chart</i>
Carta MCUSUM	Carta CUSUM Multivariada	<i>Multivariate CUSUM Chart</i>

Carta MRRC		<i>Multivariate Ridge Residual Chart</i>
DPCA	Análise de Componentes Principais Dinâmicos	<i>Dynamic Principal Component Analysis</i>
FAC	Função de Autocorrelação	<i>Autocorrelation Function</i>
FACE	Função de Autocorrelação Estimada	<i>Estimate Autocorrelation Function</i>
FACP	Função de Autocorrelação Parcial	<i>Partial Autocorrelation Function</i>
FACPE	Função de Autocorrelação Parcial Estimada	<i>Estimate Partial Autocorrelation Function</i>
LC	Linha Central	<i>Center Line</i>
LIC	Limite Inferior de Controlo	<i>Lower Control Limit</i>
LSC	Limite Superior de Controlo	<i>Upper Control Limit</i>
MSE	Erro Médio Quadrático	<i>Mean Squared Error</i>
MA (q)	Modelo de Médias Móveis de ordem q	<i>Moving Average Model</i>
MR	Amplitudes Móveis	<i>Moving Range</i>
NIPALS		<i>Nonlinear Iterative Partial Least Square</i>
PCA	Análise de Componentes Principais	<i>Principal Component Analysis</i>
PLS	Regressão pelos Quadrados Mínimos Parciais	<i>Partial Least Square</i>
SDRL	Desvio padrão da distribuição de “Run Length”	<i>Standard Deviation of Run Length Distribution</i>
SPC	Controlo Estatístico do Processo	<i>Statistical Process Control</i>

## Símbolos

$A$	Número de componentes principais retidos
$c_j^{(Q)}$	Contribuição de cada variável relativamente à estatística $Q$
$c_j^{(t_i)}$	Contribuição de cada variável relativamente à estatística $T_A^2$
$C_j^{(T_A^2)}$	Contribuição total da variável relativamente à estatística $T_A^2$
$C_{R,\alpha}$	Ponto crítico
$d$	Número de ordem de diferenciação
$E$	Matriz residual
$F_{\alpha;p;m-p}$	Percentil à direita, para uma probabilidade $\alpha$ , da Distribuição de <i>Fisher</i> , com $p$ e $m - p$ grau de liberdade
$g$	Factor multiplicativo
$h$	Limite de controlo da carta MCUSUM; número de graus de liberdade da distribuição Qui-Quadrado
$H$	Limite de controlo da carta MEWMA
$I$	Matriz identidade
$m$	Número de observações recolhidas para cada variável, quando se considera o estudo de cartas multivariadas
$n$	Dimensão de cada amostra
$p$	Número de variáveis, quando se considera o estudo de cartas multivariadas; número de ordem do modelo autorregressivo
$P$	Matriz dos vectores próprios
$p_{ij}$	Elemento da $i$ -ésima linha e $j$ -ésima coluna da matriz $P$ ;
$q$	Número de ordem do modelo médias móveis
$Q$	Estatística para a determinação dos resíduos obtidos pela PCA
$T$	Matriz dos componentes principais
$T^2$	Estatística de <i>Hotelling</i>
$T_A^2$	Estatística de <i>Hotelling</i> baseada nos $A$ componentes principais
$T_{(i)}^2$	Valor da estatística de <i>Hotelling</i> calculada a partir de $(p - 1)$ variáveis, para o método da estatística $d$
$T_j^2$	Termos incondicionais obtidos da decomposição de $T^2$ em componentes ortogonais
$T_{j+1,1,2,\dots,j}^2$	Termos condicionais obtidos da decomposição de $T^2$ em componentes ortogonais
$T_t^2$	Valor da estatística da carta MEWMA, no instante $t$
$t_i$	Valor do componente principal $i$
$R$	Diagonal entre as constantes de amortecimento das $p$ variáveis, considerando a carta MEWMA
$S$	Matriz de covariância

$\mathbf{X}$	Matriz de dados composta pelas $p$ variáveis, em que cada variável tem $m$ observações, cujas variáveis possuem média zero e variância unitária
$\mathbf{X}_i$	Vector constituído pelas observações das $p$ variáveis no instante $i$
$X_{j,novo}$	Valor da observação da variável $j$
$\hat{X}_{j,novo}$	Estimativa da variável $j$ resultante do modelo PCA
$X_t$	Observações da variável no instante $t$
$Y_t^*$	Valor da estatística da carta MCUSUM, no instante $t$
$z_\alpha$	Percentil à direita, para uma probabilidade $\alpha$ , da Distribuição Normal Reduzida
$\alpha$	Nível de significância ou erro do tipo I
$\beta$	Erro do tipo II
$\beta_{\alpha, (\frac{p}{2}, \frac{m-p-1}{2})}$	Percentil à direita, para uma probabilidade $\alpha$ , da Distribuição Beta, com parâmetros $p/2$ e $(m - p - 1)/2$
$\omega$	Valor de uma determinada estatística (contexto do SPC)
$\boldsymbol{\mu}$	Vector das médias
$\boldsymbol{\Sigma}$	Matriz das covariâncias
$l$	Atraso incorporado na matriz de dados na DPCA
$\chi^2$	Estatística do Qui-Quadrado
$\varepsilon_t$	Ruído branco
$\delta$	Alteração no parâmetro médio do processo
$\delta^2(\boldsymbol{\mu})$	Parâmetro de não centralidade
$\sigma_\varepsilon$	Desvio padrão da série de ruído branco
$\sigma_\varepsilon^S$	Desvio padrão da série de ruído branco, estimado por simulação
$\lambda_i$	Valores próprios dos $i$ componentes principais
$\lambda_j$	Constante de amortecimento para variável $j$ , na carta MEWMA
$\theta$	Parâmetro de médias móveis
$\theta_i$	Valor obtido com base nos valores próprios da matriz das covariâncias dos resíduos
$\phi$	Parâmetro autorregressivo
$\phi_{kk}$	Coefficiente de autocorrelação parcial de ordem $k$
$\rho_k$	Coefficiente de correlação de desfazamento $k$ (lag $k$ )
$\gamma_k$	Autocovariância de desfazamento $k$
$\gamma_0$	Autocovariância de desfazamento $k = 0$



# ÍNDICE GERAL

<b>Agradecimentos .....</b>	<b>I</b>
<b>Resumo .....</b>	<b>II</b>
<b>Abstract .....</b>	<b>III</b>
<b>Lista de Abreviaturas e Símbolos.....</b>	<b>IV</b>
<b>ÍNDICE GERAL.....</b>	<b>VIII</b>
<b>ÍNDICES DE FIGURAS.....</b>	<b>XI</b>
<b>ÍNDICES DE TABELAS.....</b>	<b>XII</b>
<b>1. Introdução.....</b>	<b>1</b>
1.1. Tema e Motivação.....	1
1.2. Objectivos.....	2
1.3. Estrutura da Dissertação.....	3
<b>2. Fundamentos Teóricos.....</b>	<b>5</b>
2.1. Controlo Estatístico do Processo .....	5
2.1.1 Princípios das Cartas de Controlo .....	6
2.1.2 Principais Cartas de Controlo de Variáveis.....	8
2.1.3 Regras para Detecção de Causas Especiais.....	9
2.1.4 Medidas de Desempenho das Cartas de Controlo .....	9
2.2. Controlo Estatístico do Processo com Dados Autocorrelacionados .....	11
2.2.1. Metodologia de Box- Jenkins .....	12
2.2.2. Cartas de Controlo Univariadas para Dados Autocorrelacionados .....	18
2.2.3. Cartas de Controlo Multivariadas Tradicionais .....	19
2.2.2.1. Interpretação da Carta $T^2$ .....	24
2.2.4. Cartas de Controlo Multivariadas Baseadas em Métodos de Projectão .....	27
2.2.4.1. Interpretação da Carta DPCA .....	32
2.2.4.2. Cartas de Controlo Multivariadas Baseadas na Análise de Componentes Principais Dinâmicas .....	34
2.3. Síntese .....	35
<b>3. Metodologia .....</b>	<b>36</b>
3.1. Apresentação da Metodologia Proposta .....	37

3.2.	Construção do Programa de Simulação .....	38
3.3.	Comparação do Desempenho das Cartas de Controlo .....	43
3.4.	Interpretação das Cartas de Controlo Multivariadas .....	44
3.5.	Aplicação Prática .....	45
<b>4.</b>	<b>Desenvolvimento Prático .....</b>	<b>46</b>
4.1.	Identificação das Variáveis do Processo .....	46
4.2.	Análise de Sensibilidade das Cartas Através do ARL .....	47
4.2.1.	Determinação do $ARL_{EC}$ e do $SDRL_{EC}$ .....	47
4.2.2.	Introdução de Perturbações na Média do Processo .....	51
4.2.2.1.	Desempenho da Carta $T^2$ a Alterações na Média do Processo .....	52
4.2.2.2.	Desempenho da Carta DPCA a Alterações na Média do Processo .....	54
4.2.2.3.	Comparação do Desempenho das Duas Cartas .....	56
4.3.	Metodologias de Identificação de Causas Especiais em Cartas Multivariadas .....	62
4.3.1	Construção da Carta $T^2$ e Aplicação do Método da Estatística d e do Método de Decomposição em Componentes Ortogonais .....	63
4.3.2	Construção da Carta DPCA e Aplicação do Gráfico de Contribuição .....	67
4.4.	Vantagens e Desvantagens das Duas Cartas de Controlo .....	71
4.5.	Síntese .....	73
<b>5.</b>	<b>Conclusão e Sugestões para Trabalhos Futuros .....</b>	<b>75</b>
5.1.	Conclusões gerais .....	75
5.2.	Sugestões para Trabalhos Futuros .....	77
	<b>Referências Bibliográficas .....</b>	<b>78</b>
	<b>ANEXOS .....</b>	<b>82</b>
	<b>ANEXO I – Aplicação de cartas de controlo multivariadas ao processo .....</b>	<b>83</b>
I.1.	Valores de ARL e SDRL considerando alterações no parâmetro médio do processo	83
I.1.1.	Estudo do desempenho da carta $T^2$ .....	83
I.1.2.	Estudo do desempenho da carta DPCA .....	84
I.2.	Carta $T^2$ .....	85
I.3.	Carta DPCA .....	86

## ÍNDICES DE FIGURAS

Figura 2.1 – Carta de controlo.....	7
Figura 2.2 – Geração de uma série temporal $Y_t$ .....	12
Figura 2.3 – Exemplo de um gráfico de contribuição.....	32
Figura 3.1 – Fases que compõem a metodologia a adoptar .....	37
Figura 3.2 – Construção da carta $T^2$ e determinação do $ARL_{EC}$ e $SDRL_{EC}$ .....	39
Figura 3.3 – Construção da carta DPCA e determinação do $ARL_{EC}$ e $SDRL_{EC}$ .....	40
Figura 3.4 – Estudo do desempenho da carta $T^2$ a alterações no parâmetro médio .....	41
Figura 3.5 – Estudo do desempenho da carta DPCA a alterações no parâmetro médio.....	43
Figura 3.6 – Interpretação de cartas multivariadas.....	44
Figura 4.1 – Comportamento da carta $X_1$ .....	47
Figura 4.2 – Gráfico de Superfície 3D em que se varia o valor de ARL, do número de ciclos e da dimensão da série .....	49
Figura 4.3 – Gráfico de Superfície 3D em que se varia o valor de SDRL, do número de ciclos e da dimensão da série .....	49
Figura 4.4 – Tipos de Perturbação.....	52
Figura 4.5 – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo considerando o estudo da carta $T^2$ .....	53
Figura 4.6 – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo considerando o estudo da carta DPCA .....	56
Figura 4.7 – Valores de ARL e SDRL para diferentes alterações na média da variável 1.....	57
Figura 4.8 – Valores de ARL e SDRL para diferentes alterações na média da variável 1 e 2...	58
Figura 4.9 – Valores de ARL e SDRL para diferentes alterações na média das três variáveis .	59
Figura 4.10 – Carta $T^2$ .....	60
Figura 4.11 – Carta DPCA .....	61
Figura 4.12 – Carta $T^2$ para os primeiros 60 pontos .....	64
Figura 4.13 – Critério de <i>Kaiser</i> para identificação do número de componentes principais.....	68
Figura 4.14 – Carta $T_A^2$ para os primeiros 60 pontos .....	69
Figura 4.15 – Gráfico dos componentes principais normalizados .....	70
Figura 4.16 – Gráfico de contribuição das variáveis .....	70

## ÍNDICES DE TABELAS

Tabela 2.1 – Caracterização dos correlogramas para os três processos .....	14
Tabela 2.2 – Decomposição de $T^2$ em componentes ortogonais para $p=3$ .....	26
Tabela 4.1 – Valores de $ARL_{EC}$ e $SDRL_{EC}$ para diferentes números de ciclos e dimensão da série .....	48
Tabela 4.2 – Determinação dos valores de $ARL_{EC}$ e respectivo $SDRL_{EC}$ para a carta $T^2$ e $T_A^2$ alterando o nível de significância .....	50
Tabela 4.3 – Valores de $ARL_{EC}$ e de $SDRL_{EC}$ para as duas cartas .....	51
Tabela 4.4 – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo considerando o estudo da carta $T^2$ .....	53
Tabela 4.5 – Valores de ARL e SDRL para diferentes alterações no parâmetro médio do processo quando se introduz uma perturbação na variável 1, considerando o estudo da carta DPCA.....	54
Tabela 4.6 – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo, considerando o estudo da carta DPCA .....	55
Tabela 4.7 – Valores de ARL e SDRL para diferentes alterações na média da variável 1 .....	57
Tabela 4.8 – Valores de ARL e SDRL para diferentes alterações na média da variável 1 e 2 ..	58
Tabela 4.9 – Valores de ARL e SDRL para diferentes alterações na média das três variáveis	59
Tabela 4.10 – Valores de $s_{ij}$ da matriz de covariância ( $S$ ) .....	63
Tabela 4.11 – Estatística $d$ .....	65
Tabela 4.12 – Decomposição em componentes ortogonais .....	66
Tabela 4.13 – Valores próprios correspondentes aos três componentes principais retidos e percentagem de variância cumulativa dos mesmos .....	68
Tabela 4.14 – Matriz dos vectores próprios ( $P$ ) .....	69

## 1. Introdução

Neste capítulo, pretende-se realizar uma apresentação do tema abordado nesta dissertação, as principais motivações que levaram ao desenvolvimento desta investigação bem como os principais objectivos que se pretende obter. Por fim, apresenta-se a estrutura desta dissertação.

### 1.1. Tema e Motivação

A maioria das indústrias tem como objectivo a melhoria contínua da qualidade dos seus produtos uma vez que a qualidade passou a ser um factor de decisão, fundamental, na escolha de produtos e serviços pelos clientes. Para que se consiga atingir este objectivo, é necessário o uso de técnicas estatísticas convenientes para que os processos produtivos sejam estáveis e que funcionem com um mínimo de variabilidade. Neste contexto, o controlo estatístico do processo (SPC – *Statistical Process Control*), através da implementação de cartas de controlo, permite monitorizar o desempenho do processo ao longo do tempo, detectar a ocorrência de causas especiais de variação e, por conseguinte, permitir alcançar uma redução da variabilidade da característica da qualidade do processo.

Com os meios informáticos disponíveis nos processos industriais modernos, que permitem a medição simultânea e em tempo real de várias variáveis, as indústrias têm sentido necessidade de recorrer a técnicas multivariadas, de forma a conseguirem controlar correctamente várias variáveis em simultâneo. O grande desafio que se coloca hoje em dia às empresas que pretendam adoptar estas técnicas prende-se com o facto de, em muitas situações, os dados a controlar exibirem autocorrelação (uma observação num dado instante depende de observações ocorridas em instantes anteriores).

Deste modo, o controlo estatístico multivariado do processo com dados autocorrelacionados, tem-se revelado extremamente útil, para as indústrias com processos modernos isto é, com processo contínuos, processos com monitorização automatizada ou com sistemas de aquisição de dados *on-line*.

De forma a dar resposta a esta necessidade, a presente dissertação centrou-se na realização de um estudo comparativo do desempenho de duas cartas de controlo multivariadas, com dados autocorrelacionados, nomeadamente a carta  $T^2$  de *Hotteling*, desenvolvida em 1947, e a carta DPCA (*Dynamic Principal Component Analysis*), desenvolvida por Ku *et al.* (1995). A carta  $T^2$  é a mais conhecida na literatura e, como tal, a mais recomendada para processos que

apresentam várias características da qualidade, podendo ser aplicada, também, a dados autocorrelacionados. Em alternativa, escolheu-se uma carta de controlo baseada em métodos de projecção, mais concretamente baseada na análise de componentes principais dinâmica, por se considerar que a análise de componentes principais é uma ferramenta poderosa na monitorização de processos complexos, permitindo assim reduzir significativamente o número de variáveis a serem monitorizadas e também poder ser aplicada a dados autocorrelacionados.

Os dados que permitiram a realização da componente prática, foram retirados de uma indústria química (processo contínuo) onde impera a presença de autocorrelação nos dados.

Considerou-se que este trabalho de investigação seria uma mais valia no sentido de poder-se comparar o desempenho de duas cartas multivariadas, assentes em princípios diferentes, procurando saber qual apresenta o melhor desempenho quando o processo é sujeito a perturbações ao nível da média numa ou mais variáveis. Por outro lado, por existir na literatura poucos estudos desta natureza, ou seja, estudos que analisem o desempenho da carta DPCA através do ARL (*Average Run Length*) e a comparem com outro tipo de cartas multivariadas, recorrendo à simulação. E também identificar as principais vantagens e desvantagens que cada uma das cartas pode apresentar na óptica do utilizador.

## 1.2. Objectivos

A presente dissertação centra-se num estudo comparativo do desempenho de duas cartas de controlo multivariadas com dados autocorrelacionados, nomeadamente a carta  $T^2$  de *Hotelling*, desenvolvida em 1947 e a carta DPCA, baseada na análise de componentes principais dinâmica de *Ku et al.* (1995).

Deste modo, apresentam-se os objectivos considerados mais relevantes, que se pretendem atingir com a presente dissertação:

- ✓ Objectivo principal: desenvolver uma metodologia que permita estabelecer uma comparação entre o desempenho da carta  $T^2$  com a carta DPCA, baseada na análise de componentes principais dinâmica, de forma a indicar qual delas é a mais adequada, quando o processo é sujeito a um tipo de perturbação ao nível da média, numa ou mais variáveis em simultâneo.

- ✓ Proceder a uma revisão bibliográfica do controlo estatístico do processo, nomeadamente, cartas de controlo univariadas e multivariadas, processos com dados autocorrelacionados, interpretação das cartas de controlo multivariadas.
- ✓ Desenvolver detalhadamente os fundamentos teóricos relativos à carta  $T^2$  e à carta DPCA, baseada na análise de componentes principais dinâmica. Apresentar os métodos considerados mais adequados para a interpretação das causas especiais.
- ✓ Proporcionar informações claras e precisas sobre os procedimentos que devem ser seguidos para aplicação das cartas de controlo multivariadas quando se está perante processos com dados autocorrelacionados.
- ✓ Terminado o trabalho de investigação, pretende-se identificar novas áreas de investigação, neste domínio.

### 1.3. Estrutura da Dissertação

A presente dissertação está estruturada em 5 capítulos, descrevendo-se cada um deles em seguida:

No Capítulo 1, designado como *Introdução*, é feita uma introdução ao tema abordado na dissertação, “análise do desempenho de cartas de controlo multivariadas com dados autocorrelacionados”, evidenciando-se a motivação e relevância do seu estudo. Seguidamente definem-se os objectivos e apresenta-se a estrutura da dissertação no qual são descritos os aspectos fundamentais abordados em cada capítulo.

No Capítulo 2, denominado como *Fundamentos teóricos*, é feita uma breve introdução ao controlo estatístico do processo, onde se refere, resumidamente, os princípios das cartas de controlo, as principais cartas de controlos, as regras para detecção de causas especiais e as medidas de desempenho destas mesmas cartas. O controlo estatístico do processo com dados autocorrelacionados é igualmente abordado onde se apresenta a metodologia de *Box Jenkins* (modelos ARIMA) e um conjunto de cartas de controlo (univariadas e multivariadas) concebidas para este tipo de processos, dando-se maior relevância às cartas multivariadas. Relativamente às cartas multivariadas, dá-se especial atenção à carta  $T^2$  e à carta baseada na análise de componentes principais dinâmica (DPCA). Apresentam-se, também, os métodos para a interpretação de causas especiais nestas duas cartas.

No Capítulo 3, denominado por *Metodologia*, apresenta-se a metodologia que se sugere para a realização do estudo comparativo das duas cartas de controlo multivariadas com dados

autocorrelacionados. Esta metodologia baseia-se nos desenvolvimentos apresentados no Capítulo 2 e foi elaborada recorrendo a fluxogramas.

No Capítulo 4, *Desenvolvimento prático*, é realizada a parte experimental deste trabalho. Apresentam-se as variáveis relevantes para a monitorização do processo, de forma a indicar qual o tipo de cartas a aplicar ao processo. Realizam-se dois estudos, no primeiro estudo é feita uma análise comparativa do desempenho da carta  $T^2$  e da carta baseada na análise de componentes principais dinâmica (DPCA), através de simulação, quando o processo é sujeito a perturbações numa ou mais variáveis em simultâneo. No segundo apresentam-se três métodos diferentes para a interpretação das duas cartas multivariadas, quando o processo é sujeito a perturbações numa variável. Por fim apresenta-se as conclusões que se retiraram destes dois estudos.

No Capítulo 5, designado como *Conclusões e Sugestões para Trabalhos Futuros*, apresenta-se as conclusões gerais do trabalho realizado no âmbito desta dissertação e sugestões para o desenvolvimento de trabalhos futuros.



## 2. Fundamentos Teóricos

Tendo em conta o tipo de processo sobre o qual incide a presente dissertação, este capítulo foca, fundamentalmente, o controlo estatístico do processo com dados autocorrelacionados, dando maior importância as cartas de controlo multivariadas concebidas para este tipo de processos.

Relativamente às cartas multivariadas desenvolvidas para processos com dados autocorrelacionados, dá-se especial atenção à carta  $T^2$  e à carta baseada na análise de componentes principais dinâmica (DPCA). Apresentam-se, também, os métodos para a interpretação destas duas cartas.

### 2.1. Controlo Estatístico do Processo

O controlo estatístico do processo (adiante designado por SPC), desenvolvido em 1924 por *Walter Shewhart* permite monitorizar o desempenho do processo ao longo do tempo, e detectar eventos incomuns que possam ocorrer. Com a utilização desta técnica estatística, é possível descrever detalhadamente o comportamento do processo, identificar a sua variabilidade e possibilitar o seu controlo, através da recolha contínua de dados e da análise de possíveis causas de variação que possam afectar o processo.

Esta técnica é extremamente importante para qualquer indústria uma vez que a sua aplicação é relativamente simples, podendo ficar a cargo do próprio operador de um equipamento ou do realizador de um serviço. Além disso, a velocidade com que se consegue detectar mudanças no comportamento do processo, permite a tomada de acções correctivas atempadamente, evitando assim maiores prejuízos.

No entanto, qualquer produto ou serviço, está sujeito a uma variabilidade inevitável, ou seja, existem várias fontes de variação que podem afectar um processo produtivo, como, por exemplo, condições dos equipamentos, condições ambientais, métodos de trabalho, matérias-primas, instrumentos de medição e por fim o factor humano envolvido no processo.

Relativamente às fontes de variação que podem afectar um processo, *Shewhart* classificou-as em dois tipos:

- Causas Comuns são causas aleatórias, ou seja, os valores individuais de uma determinada característica são diferentes mas o seu conjunto segue um certo padrão que pode ser descrito por uma distribuição de probabilidade caracterizada por uma

determinada forma e por parâmetros de localização e de dispersão (Pereira e Requeijo, 2008).

- Causas Especiais são causas esporádicas que não se inserem na distribuição seguida por uma característica quando o processo está sob controlo estatístico (Pereira e Requeijo, 2008).

As causas comuns, são inevitáveis, inerentes ao próprio processo e, caso sejam identificadas, devem ser reduzidas ao máximo uma vez que não é possível eliminá-las na sua totalidade. Estas causas, geralmente, estão relacionadas com cinco factores, nomeadamente, a mão-de-obra, a matéria-prima, o meio ambiente, as máquinas e os métodos. Já as causas especiais, normalmente provêm de erros de operadores, de alterações ocasionais de condições de trabalho, ajustes inadequados das máquinas, entre outros. Estas causas, caso sejam detectadas, devem ser eliminadas e sugeridas melhorias no processo e na qualidade dos produtos. Para que se consiga fazer esta análise, isto é, detectar as causas especiais de variação e as causas comuns, caso apareçam, deve-se recorrer à monitorização através das cartas de controlo.

Se um processo está sujeito apenas a causas comuns de variação, então diz-se que este está sob controlo estatístico, ou seja, que apresenta uma variabilidade estatisticamente estável e uniforme, em termos de média e dispersão, ao longo do tempo. Nestas condições, é possível estimar os parâmetros do processo e determinar a sua capacidade de forma a produzir de acordo com as especificações estabelecidas. Por outro lado, se um processo exhibe causas especiais de variação, diz-se que está fora de controlo estatístico, apresentando uma variabilidade estatisticamente instável, em termos de média e dispersão, ao longo do tempo.

### 2.1.1 Princípios das Cartas de Controlo

As cartas de controlo são a principal ferramenta utilizada no SPC pois servem de apoio ao controlo da qualidade de um processo, reduzindo a quantidade de produtos fora das especificações e, consequentemente, os custos de produção. Estas cartas quando são usadas adequadamente, permitem ao operador um controlo contínuo do processo, levando a que se produza de forma consistente, com qualidade e com custos adequados.

Uma carta de controlo, é definida por um gráfico que mostra a evolução ao longo do tempo de uma estatística ( $\omega$ ) para uma determinada característica da qualidade. Para além dos valores da estatística, também é representada a linha central (LC), e os limites superior (LSC) e inferior (LIC) de controlo. A Figura 2.1 apresenta a forma geral de uma carta de controlo.

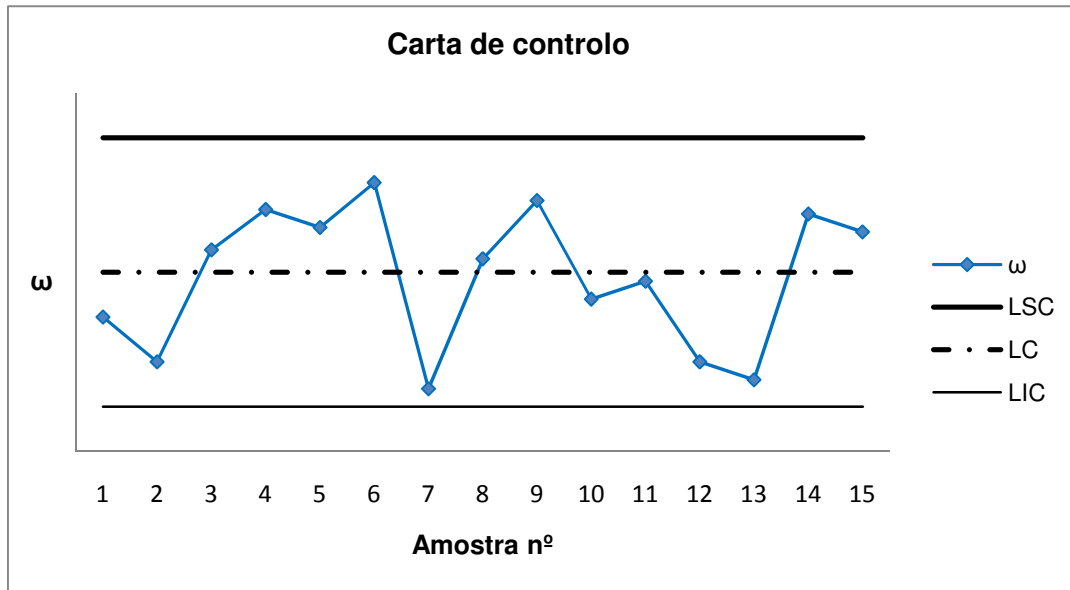


Figura 2.1 – Carta de controle

Se todos os valores da estatística estiverem dentro dos limites de controle, sem qualquer tendência e a disposição dos pontos dentro dos limites for aleatória, então considera-se que o processo está sob controle estatístico. Caso haja um ou mais pontos fora dos limites de controle ou se estes apresentarem uma tendência especial ou sistemática, mesmo que os pontos estejam dentro dos limites, considera-se que o processo está fora de controle estatístico.

Segundo *Shewhart*, considerando que os valores da estatística  $\omega$  seguem uma distribuição aproximadamente Normal  $N(\mu_\omega, \sigma_\omega^2)$ , os limites de controle e a linha central são dados por:

$$LSC_\omega = \mu_\omega + 3\sigma_\omega$$

$$LC_\omega = \mu_\omega \quad (2.1)$$

$$LIC_\omega = \mu_\omega - 3\sigma_\omega$$

Como se pode verificar, estes limites estão localizados a  $\pm 3\sigma_\omega$  da linha central o que significa, considerando que  $\omega$  é Normalmente distribuído, que a probabilidade de um ponto estar entre os limites de controle é de 99,73%. Assim, os restantes 0,27% representam a probabilidade de um ponto ainda pertencente à distribuição de  $\omega$  estar fora dos limites de controle, designando-se por falso alarme. A esta probabilidade, dá-se o nome de risco do erro tipo I, risco do produtor ou nível de significância. Existe ainda o erro tipo II, que é designado também por risco do consumidor, que corresponde a considerar o processo sob controle estatístico quando, na verdade, está fora de controle.

Quando se pretende construir uma carta de controlo, é necessário distinguir-se duas fases na sua implementação. A Fase 1, que corresponde a um controlo retrospectivo do processo no qual se verifica a estabilidade do mesmo, estimam-se os seus parâmetros ( $\mu, \sigma^2$ ) e fixam-se os limites de controlo. A Fase 2, que corresponde à monitorização do processo, onde se procura identificar possíveis causas especiais de variação e caso estas existam, eliminá-las de imediato. Para uma análise mais profunda dos principais passos a seguir na aplicação de cada uma das Fases (1 e 2), podem citar-se Pereira e Requeijo (2008).

As cartas de controlo são normalmente classificadas de acordo com o tipo de característica da qualidade que elas monitorizam. Assim, é necessário identificar-se o tipo de característica que se pretende controlar:

- Variáveis são todas as características que podem ser expressas como valores numa escala contínua de medidas (dimensões, peso, temperatura). As cartas de controlo que representam estas características são designadas de cartas de controlo de variáveis.
- Atributos são todas as características que não podem ser expressas numa escala contínua assumindo apenas valores discretos (como por exemplo, o número de artigos defeituoso, etc). As cartas de controlo que representam estas características são designadas de cartas de controlo de atributos.

Por sua vez, as cartas de controlo também podem ser classificadas de acordo com o número de características da qualidade monitorizadas num processo:

- Cartas de controlo univariadas são usadas quando se pretende controlar apenas uma característica da qualidade.
- Cartas de controlo multivariadas são usadas quando se pretende controlar mais que uma característica de qualidade num mesmo gráfico.

### 2.1.2 Principais Cartas de Controlo de Variáveis

As cartas de controlo de variáveis são utilizadas para analisar e ajustar a variação de um processo em função do tempo. Estas cartas, de um modo geral, fornecem mais informação sobre o processo de produção do que as cartas de controlo de atributos, tornando-se assim mais eficientes.

As cartas de variáveis mais conhecidas são as cartas de *Shewhart*, CUSUM e EWMA. Relativamente às cartas de *Shewhart*, as mais utilizadas são as cartas da média e da amplitude ( $\bar{X} - R$ ) ou desvio padrão ( $\bar{X} - S$ ), no qual é obrigatória a recolha de amostras com

duas ou mais observações ( $n \geq 2$ ), e as cartas de valores individuais e amplitudes móveis ( $X - MR$ ). Estas cartas apenas podem ser aplicadas a variáveis independentes e normalmente distribuídas. Se a distribuição dos dados do processo não for normal, então deve-se ajustá-la de forma a que seja, para que as considerações sobre as cartas de controlo de variáveis sejam válidas.

A maior desvantagem destas cartas de controlo de *Shewhart* é que são insensíveis a pequenas variações nos processos e também porque utilizam apenas a informação do último ponto. Deste modo, quando se pretende detectar alterações pequenas e moderadas dos parâmetros do processo (da ordem de  $1,5\sigma$  ou inferiores), utilizam-se as cartas de controlo CUSUM (*Cumulative Sum*) e EWMA (*Exponentially Weighted Moving Average*). As cartas CUSUM foram introduzidas por *Page*, em 1954, mas também receberam a atenção de outros autores, tanto para o controlo da média como para o controlo da variância. As cartas EWMA foram introduzidas por *Roberts*, em 1959, e mais tarde foram objecto de estudo de outros autores, tanto para o controlo da média como para o controlo da variância.

### 2.1.3 Regras para Detecção de Causas Especiais

Como se referiu no ponto 2.1.1, quando um ponto está fora dos limites de controlo ou quando estes apresentam um padrão sistemático, considera-se que se está na presença de causas especiais de variação.

Para que se consiga identificar claramente, nas cartas de controlo de *Shewhart*, a presença de causas especiais de variação tem-se por base oito regras propostas por *Western Electric Handbook* sendo estas incluídas na norma ISO 8258:1991.

Na presente dissertação apenas se vai utilizar a primeira regra para identificar a presença de causas especiais nas duas cartas de controlo multivariadas. Assim, o processo encontra-se fora de controlo estatístico quando se verifica “um qualquer ponto fora dos limites de acção (limites  $\pm 3\sigma_\omega$ )”.

### 2.1.4 Medidas de Desempenho das Cartas de Controlo

A distribuição “*Run Length*” é utilizada para descrever o desempenho de uma carta de controlo. O termo “*Run Length*” representa o número de observações recolhidas até obter uma observação fora dos limites de controlo, dada pela alteração da média. O valor médio da distribuição “*Run Length*” é designado por ARL (*Average Run Length*) e o valor correspondente

ao desvio padrão por SDRL (*Standard Deviation Run Length*). Por sua vez, o ARL é considerado como a medida de desempenho mais usada das cartas de controlo.

Pretende-se que o valor do ARL seja o maior possível quando o processo se encontra sob controlo estatístico (adiante designado por  $ARL_{EC}$ ), de forma a minimizar o número de falsos alarmes. Mas para que seja possível detectar rapidamente uma alteração do parâmetro do processo, o valor do ARL deverá ser o menor possível quando o processo se encontra fora de controlo (adiante designado por  $ARL_{FC}$ ). Quando se realizam comparações entre cartas de controlo, aquela que apresentar em simultâneo o maior valor de ( $ARL_{EC}$ ), e o menor valor ( $ARL_{FC}$ ) deverá ser considerada como a mais eficaz.

Numa carta de controlo de *Shewhart*, o ARL é determinado em função de  $\alpha$  (erro tipo I) quando o processo está sob controlo e  $\beta$  (erro tipo II) quando o processo está fora de controlo. Assim, o ARL e o seu respectivo desvio padrão (SDRL), em controlo e fora de controlo, são dados pelas seguintes expressões (Pereira e Requeijo, 2008):

$$ARL_{Em\ Controlo} = ARL_{EC} = \frac{1}{\alpha} \quad (2.2)$$

$$SDRL_{Em\ Controlo} = SDRL_{EC} = \frac{\sqrt{1-\alpha}}{\alpha} \quad (2.3)$$

$$ARL_{Fora\ de\ Controlo} = ARL_{FC} = \frac{1}{1-\beta} \quad (2.4)$$

$$SDRL_{Fora\ de\ Controlo} = SDRL_{FC} = \frac{\sqrt{\beta}}{1-\beta} \quad (2.5)$$

Como se viu no ponto 2.1.1, a probabilidade da carta de controlo indicar incorrectamente uma alteração na média do processo, isto é, de ocorrer um falso alarme é de 0,27%. Assim, substituindo essa probabilidade na expressão 2.2, obtém-se um  $ARL_{EC}$  igual a 370. Isto significa que, estando o processo sob controlo, de 370 em 370 pontos existirá, em média, um falso alarme.

O ARL é muito utilizado para comparar o desempenho de várias cartas (univariadas e multivariadas). Para tal, devem ajustar-se os limites de controlo para obter o mesmo valor de ARL sob controlo estatístico de modo a estabelecer-se uma comparação de ARL's para várias dimensões de desvio ao processo ( $\delta$ ).

Uma vez que a presente dissertação tem como objectivo desenvolver um estudo comparativo do desempenho de duas cartas de controlo multivariadas com dados autocorrelacionados, apresentam-se de seguida algumas referências no qual se realizaram estudos comparativos de cartas de controlo multivariadas.

O artigo desenvolvido por *Lowry et al.* (1992) apresenta um estudo comparativo de 4 cartas de controlo multivariadas, nomeadamente, a carta  $\chi^2$  de *Hotelling*, a carta MCUSUM desenvolvida por *Crosier* (1988), a carta multivariada CUSUM (MC1) desenvolvida por *Pignatiello e Runger*, em 1990, a carta MEWMA no qual se utiliza a matriz covariância desenvolvida por *Tsui e Woodall*, em 1991 e a carta MEWMA no qual se utiliza a matriz covariância desenvolvida por *MacGregor e Harris*, em 1990, tendo como medida de desempenho o ARL.

O artigo desenvolvido por *Javaheri e Houshmand* (2001) apresenta um estudo comparativo do desempenho de cinco métodos de controlo da qualidade multivariada, nomeadamente, a carta  $T^2$  desenvolvida por *Hotelling*, em 1947, a carta multivariada de *Shewhart* desenvolvida por *Houshmand et al.*, em 1997, a Análise Discriminante desenvolvida por *Murphy*, em 1987, a Decomposição de  $T^2$  desenvolvida por *Mason, Tracy e Young* (1995) e a carta MRRC (*Multivariate Ridge Residual*) desenvolvida por *Houshmand e Javaheri*, em 1997. Este estudo é realizado através a metodologia de simulação Monte Carlo e utiliza como medida de desempenho o ARL.

O artigo desenvolvido por *Ghute e Shirke* (2008) apresenta a carta  $T^2$  sintética e compara o desempenho desta carta com a carta  $T^2$  desenvolvida por *Hotelling*, em 1947 e com a carta  $T^2$  com regras de funcionamento suplementares desenvolvida por *Aparisi et al.* (2004), através do ARL. Os valores de ARL das cartas  $T^2$  sintética e  $T^2$  são determinados através do desenvolvimento de um programa computacional usando o *software* MATLAB. Já os valores da carta  $T^2$  com regras suplementares são dados por *Apirisi et al.* (2004).

## 2.2. Controlo Estatístico do Processo com Dados Autocorrelacionados

Quando se utilizam as cartas de controlo (univariadas e multivariadas), assume-se que os dados relativos à característica da qualidade são independentes e identicamente distribuídos segundo uma distribuição Normal com média ( $\mu$ ) e variância ( $\sigma^2$ ). No entanto, na prática, muitas vezes a suposição de independência é violada pelo facto das variáveis serem autocorrelacionadas, isto é, o valor da variável num dado instante depende, em parte, do(s) valor(es) anterior (es) da mesma variável. Desta forma, as cartas de controlo não funcionam correctamente pois se for assumida a independência dos dados, quando na realidade existe a presença de autocorrelação, tanto os limites de controlo como os parâmetros do processo serão estimados incorrectamente. Quando a autocorrelação é positiva, o desvio padrão é subestimado, o que resulta num estreitamento dos limites de controlo originando um aumento do número de falsos alarmes. Quando a autocorrelação é negativa, o desvio padrão é

sobrestimado, produzindo um efeito oposto, isto é, aumento dos limites de controlo seguido duma perda de sensibilidade às variações da média do processo (Wieringer, 1999).

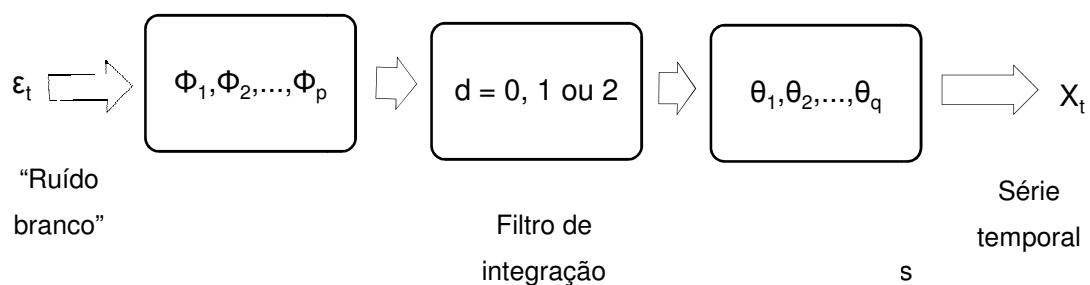
Assim, quando se pretende implementar o SPC nos processos industriais, deve-se primeiramente verificar os pressupostos subjacentes à implementação de cartas de controlo (verificar se os dados são independentes e normalmente distribuídos) e só depois controlá-los estatisticamente.

Sempre que o pressuposto da independência dos dados é violada, ou seja, os dados exibem autocorrelação, sugere-se a aplicação da metodologia de *Box* e *Jenkins* para modelar correctamente o comportamento dinâmico dos dados.

### 2.2.1. Metodologia de *Box-Jenkins*

A metodologia de *Box-Jenkins* é uma ferramenta importante de fácil aplicação para previsão de variáveis baseadas em séries temporais. Esta metodologia baseia-se no ajuste de modelos matemáticos, denominados por ARIMA (*Auto Regressive Integrated Moving Average*) às observações recolhidas, de forma que a diferença entre os valores gerados pelos modelos e os valores observados resulte em séries de resíduos, de comportamento aleatório em torno de zero (Miranda, 2002).

Os modelos ARIMA resultam da combinação de três componentes designados “filtros”: o componente auto-regressivo (AR), o filtro de integração (I) e o componente de médias móveis (MA). Uma série pode ser modelada pelos três filtros ou apenas um subconjunto deles, resultando assim em várias alternativas de modelos.



**Figura 2.2** – Geração de uma série temporal  $Y_t$

Esta metodologia proposta por *Box* e *Jenkins* e descrita na sua obra mais recente, *Box et al.* (2008), pressupõe três etapas principais: (1) identificação do modelo, (2) estimação dos parâmetros e (3) avaliação do modelo. Apresenta-se de seguida a descrição de cada uma das etapas.



### 1) Identificação do modelo

Esta etapa consiste em descobrir qual o modelo ARIMA  $(p, d, q)$  que melhor descreve o comportamento do processo. Para tal, é necessário determinar quais dos filtros (AR, I e MA) compõem o processo gerador da série bem como os valores de  $p$ ,  $d$  e  $q$  que indicam as respectivas ordens dos filtros.

A identificação do modelo ARIMA em causa, é realizada com base na função de autocorrelação (FAC) e função de autocorrelação parcial (FACP). Através dos correlogramas resultantes destas duas funções, consegue-se indicar qual o modelo a ser utilizado.

#### Função de Autocorrelação (FAC)

Quando o valor de uma determinada variável num dado instante depende em parte do(s) valor(es) anterior(es) da mesma variável, diz-se que esta variável é autocorrelacionada. Esta autocorrelação é definida pelo coeficiente de correlação  $\rho_k$ , que mede a correlação entre observações desfasadas  $k$  períodos de tempo.

$$\rho_k = \frac{Cov(X_t - X_{t+k})}{Var(X)} = \frac{\gamma_k}{\gamma_0}, \quad k = 0, 1, 2, \dots \quad (2.6)$$

Assim  $\rho_1$  indica como é que valores sucessivos de  $X_t$  se relacionam,  $\rho_2$  indica como os valores a cada dois períodos se relacionam e assim sucessivamente.

Na equação anterior, adopta-se a seguinte simbologia:

$\gamma_k$  – autocovariância de desfasamentos  $k$ ;

$\gamma_0$  – autocovariância de desfasamento  $k = 0$ .

O conjunto formado pelos coeficientes de autocorrelação  $\rho_k$  para  $k = 0, 1, 2, \dots$ , denomina-se por Função de Autocorrelação (FAC). Uma vez que não se conhece a função de autocorrelação, esta deve ser estimada. Segundo *Box et al.* (2008), o valor estimado mais satisfatório de  $\rho_k$  é dado por  $r_k$ .

$$r_k = \hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}, \quad k = 0, 1, 2, \dots \quad (2.7)$$

O conjunto formado pelos coeficientes de correlação estimados  $r_k$  é designado por Função de Autocorrelação Estimada (FACE).

#### Função de Autocorrelação Parcial (FACP)

Utiliza-se a função de autocorrelação parcial quando o comportamento da função de autocorrelação dificulta a escolha do melhor modelo para descrever o processo.

A autocorrelação parcial de desfazamento  $k$  é usada para medir a correlação entre  $X_t$  e  $X_{t+k}$  quando os efeitos das observações  $(X_{t+1}, X_{t+2}, \dots, X_{t+k-1})$  são removidos. O coeficiente de autocorrelação parcial de ordem  $k$  é designado por  $\phi_{kk}$ .

Para modelos  $AR(p)$  (autorregressivos de ordem  $p$ ), a autocorrelação parcial é obtida recorrendo às equações de *Yule-Walker* (Box et al., 2008)

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (2.8)$$

Ou, recorrendo a notação matricial

$$\boldsymbol{\phi}_k = \mathbf{P}_k^{-1} \boldsymbol{\rho}_k \quad (2.9)$$

Como os coeficientes  $\phi_{kk}$  não são conhecidos, têm de ser estimados. Assim, deve-se substituir na expressão 2.8 os  $\rho_k$  pelos valores estimados  $r_k$ , de forma a obter as estimativas  $\hat{\phi}_{11}$ ,  $\hat{\phi}_{22}$ ,  $\hat{\phi}_{33}$ ,... O conjunto formado pelos coeficientes de autocorrelação parcial estimados  $\hat{\phi}_{kk}$ , denomina-se por Função de Autocorrelação Parcial Estimada (FACPE).

Nem sempre é fácil distinguir, através dos correlogramas, qual o processo que está em questão (AR, MA ou ARMA). Deste modo, apresenta-se de seguida uma tabela com os padrões de correlação que são esperados em processos estacionários (Murteira et al., 2000).

**Tabela 2.1** – Caracterização dos correlogramas para os três processos

Processo	FAC	FACP
<b>AR(p)</b>	<u>Infinita</u> – decai para zero exponencialmente ou segundo uma sinusoidal amortecida	<u>Finita</u> – decai bruscamente para zero a partir do lag $p$
<b>MA(q)</b>	<u>Finita</u> – decai bruscamente para zero a partir do lag $q$ ou segundo uma sinusoidal amortecida	<u>Infinita</u> – decai para zero exponencialmente
<b>ARMA(p, q)</b>	<u>Infinita</u> – decai para zero exponencialmente ou segundo uma sinusoidal amortecida	<u>Finita</u> – decai bruscamente para zero a partir do lag $q$ ou segundo uma sinusoidal amortecida

O melhor modelo deve ser parcimonioso, isto é, deve-se escolher o modelo que apresentar um menor conjunto de parâmetros a ser estimado.

## 2) Estimação dos parâmetros

Uma vez determinamos os valores de  $p$  e  $q$ , é necessário estimar os  $p$  parâmetros  $\phi$  (parâmetros autorregressivos) e os  $q$  parâmetros  $\theta$  (parâmetros média móvel) do modelo escolhido. Os métodos mais comuns para esta estimação são o método dos mínimos quadrados ou a máxima verosimilhança.

## 3) Avaliação do modelo

Uma vez escolhido o modelo ARIMA e estimados os seus parâmetros, deve-se verificar se o modelo em questão se ajusta aos dados da série temporal. Um dos métodos que pode ser utilizado é o teste de *Box-Ljung*, que permite identificar se o resíduo é um ruído branco, isto é, se é independente e identicamente distribuído. Caso não seja, conclui-se que o modelo escolhido não é o apropriado, sendo necessário encontrar um novo modelo que melhor descreva o processo em causa.

Para se aplicar a metodologia de *Box-Jenkins* é necessário primeiramente identificar a série e remover a não estacionaridade uma vez que os modelos introduzidos por *Box* e *Jenkins* descrevem exclusivamente séries estacionárias, ou seja, com média e variância constantes no tempo. Isto permite fixar os parâmetros do modelo válidos para previsão do futuro a partir do passado.

Assim, de seguida apresentam-se os modelos para séries estacionárias (modelos ARMA) e para séries não-estacionárias (modelos ARIMA). Relativamente aos primeiros modelos, estes são ajustados à série original. Os segundos modelos, são ajustados a séries já estacionárias transformadas pelo método das diferenças de ordem  $d$ , isto é, cujas séries originais são séries não-estacionárias.

## Processos estacionários – Modelos ARMA ( $p, q$ )

A combinação de  $p$  e  $q$  originam um grande número de modelos matemáticos para descrever séries temporais.

- **Processos autorregressivos – AR( $p$ )**

Os processos autorregressivos de ordem  $p$ , AR ( $p$ ) ou ARMA ( $p, 0$ ) podem ser apresentados pela seguinte expressão:

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.10)$$

onde

$\xi$  – parâmetro que determina a média do processo;

$\phi_i$  – parâmetro da componente autorregressiva, com  $i = 1, 2, \dots, p$ ;

$X_{t-p}$  – valor da variável no instante  $t - p$ ;

$\varepsilon_t$  – valor do resíduo no instante  $t$ .

Por exemplo, para um processo autorregressivo de primeira ordem AR(1), este pode ser descrito da seguinte forma:

$$X_t = \xi + \phi_1 X_{t-1} + \varepsilon_t \quad (2.11)$$

Em que a média do processo AR(1) é dada por

$$\mu = \frac{\xi}{1 - \phi_1} \quad (2.12)$$

e a função de autocorrelação (FAC) por

$$\rho_k = \phi_1^k, \text{ com } k = 0, 1, \dots \quad (2.13)$$

- **Processos de Média Móvel – MA( $q$ )**

Os processos de média móvel de ordem  $q$ , MA( $q$ ) ou ARMA( $q, 0$ ) podem ser descritos através da seguinte expressão:

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.14)$$

onde

$\mu$  – média do processo;

$\theta_i$  – parâmetro da componente média móvel, com  $i = 1, 2, \dots, q$ ;

$\varepsilon_{t-q}$  – valor do resíduo, no instante  $t - q$ .

Se se estiver perante o processo de média móvel de primeira ordem  $MA(1)$ , este pode ser representado pela seguinte forma:

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (2.15)$$

Em que a função de autocorrelação (FAC) é dada por

$$\rho_k = \begin{cases} \frac{-\theta}{1 + \theta^2} & k = 1 \\ 0 & k > 0 \end{cases} \quad (2.16)$$

- **Processos autorregressivos de Média Móvel –  $ARMA(p, q)$**

Os processos autorregressivos de média móvel,  $ARMA(p, q)$ , podem ser apresentados pela seguinte expressão

$$X_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.17)$$

Se estivermos perante um processo  $ARMA(1, 1)$ , este pode ser representado da seguinte forma

$$X_t = \xi + \phi_1 X_{t-1} - \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (2.18)$$

Em que a média é dada pela expressão (2.12) e a função de autocorrelação por

$$\rho_k = \begin{cases} \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\theta_1 \phi_1} & k = 1 \\ \phi_1 \rho_{k-1} & k > 1 \end{cases} \quad (2.19)$$

### **Processos não-estacionários – Modelos ARIMA ( $p, d, q$ )**

Como foi referido anteriormente, quando as séries são não-estacionárias, estas têm de se tornar estacionárias através do método das diferenças de ordem  $d$ . O número de diferenças ( $d$ ) necessário para tornar a série estacionária é designado de ordem de integração. Por exemplo, se a série original  $X_t$  não é estacionária, mas  $Y_t = \Delta X_t = X_t - X_{t-1}$  é estacionária, diz-se que  $X_t$  é integrada de primeira ordem ( $d = 1$ ). A ordem de integração da série depende do número de diferenças que esta necessita para se tornar estacionária. Deste modo, se a série original ( $X_t$ ) se tornar estacionária e a série resultante ( $Y_t$ ) for representada por um modelo  $ARMA(p, q)$ , diz-se que  $X_t$  é descrita por um modelo  $ARIMA(p, d, q)$  representada por

$$Y_t = \xi + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.20)$$

Para este tipo de processos, existem, também, os modelos  $ARI(p, d)$  ou  $ARIMA(p, d, 0)$  e  $IMA(d, q)$  ou  $ARIMA(0, d, q)$ .

### 2.2.2. Cartas de Controlo Univariadas para Dados Autocorrelacionados

Quando existe a presença de autocorrelação nos dados, as cartas de controlo referenciadas no ponto 2.1.3 não devem ser aplicadas para estes casos.

Deste modo, *Montgomery* (2005) sugere, para a monitorização de processos cujos dados apresentam autocorrelação, duas abordagens para a construção de cartas de controlo. A primeira abordagem utiliza as cartas de controlo tradicionais com limites de controlo modificados, considerando a autocorrelação que se verifica no processo. A segunda consiste em determinar um modelo matemático ARIMA (*Autoregressive Integrated Moving Average*) que se ajuste aos dados autocorrelacionados e construir as cartas de controlo com os resíduos ou dos erros de previsão do modelo ajustado.

Relativamente à primeira abordagem, apresentam-se alguns dos trabalhos mais relevantes, que tiverem por base a modificação dos limites das cartas de controlo. *Vasilopoulos e Stamboulis* (1978) introduziram as cartas de controlo de limites modificados para processos autorregressivos de segunda ordem, AR(2). *Yang e Hancock* (1990) desenvolveram limites modificados para as cartas  $\bar{X}$ ,  $R$ ,  $S$  e  $S^2$ , com base na função de autocorrelação  $\rho_k$ . *Montgomery e Mastrangelo* (1991) sugeriram o uso da estatística EWMA como um valor de previsão "um passo adiante", de forma a eliminar a autocorrelação dos dados. *Montgomery e Mastrangelo* (1991), também desenvolveram a carta MCEWMA (*Moving Center-line EWMA*) com limites não constantes, que permite analisar, simultaneamente, a evolução do comportamento do processo e detectar causas especiais de variação. *Zhang* (1998) propôs o uso da estatística EWMA para processos estacionários (carta EWMAST), mas ajustando a variação do processo segundo os efeitos de autocorrelação.

Já em relação à segunda abordagem, que consiste em ajustar o melhor modelo matemático ARIMA aos dados recolhidos, esta, relativamente à primeira, teve uma maior aceitação, tendo sido investigada por diversos autores.

Uma das referências mais relevantes é o trabalho desenvolvido por *Alwan e Roberts* (1988), que propõem a carta CCC (*Common-Cause Chart*) e a carta SCC (*Special-Cause Chart*). A carta CCC consiste, simplesmente, na construção dos valores previstos obtidos pelo ajuste dos dados através do modelo ARIMA conhecido por *Box e Jenkins*. Esta carta não contempla a existência de limites de controlo pelo que apenas é usada para estimar o nível actual do processo. A carta SCC, é uma carta de resíduos, que tem por base as cartas de *Shewhart*. Os dois autores, mostraram que esta carta é muito mais vantajosa que a carta de *Shewhart* tradicional no que diz respeito à alterações da média do processo. Esta carta, foi investigada por vários autores nomeadamente, *Harris e Ross* (1991), *Wardell et al.* (1994), *Runger e*

*Willemain* (1995), *Lin e Adams* (1996), *Vander Wiel* (1996a), *Reynolds e Lu* (1997), *Zhang* (1998) e *English et al.* (2000).

Outras cartas que podem ser utilizadas para monitorizar os resíduos são as cartas de controlo CUSUM e EWMA, que permitem detectar rapidamente alterações pequenas e moderadas na média do processo. Foram vários os autores que apresentaram desenvolvimentos para obtenção destas duas cartas: *Harris e Ross* (1991), *Yashchin* (1993), *Runger et al.* (1995) e *Lu e Reynolds* (2001) para carta CUSUM e *Reynolds e Lu* (1997), *Lu e Reynolds* (1999a, 1999b) para a carta EWMA.

As cartas apresentadas neste ponto não serão desenvolvidas, uma vez que na presente dissertação apenas serão utilizadas cartas de controlo multivariadas para dados autocorrelacionados.

Um outro método que pode ser utilizado na monitorização de processos cujos dados apresentam autocorrelação, é a transformação de dados, como por exemplo, as transformadas de fourier.

### 2.2.3. Cartas de Controlo Multivariadas Tradicionais

Na maior parte das indústrias, quando se tem uma única característica para controlar a qualidade de um produto, utilizam-se cartas de controlo univariadas tradicionais. Mas quando um produto é identificado por várias características da qualidade, e estas são controladas separadamente, a sua interpretação e diagnóstico pode tornar-se extremamente difícil e por vezes enganadora. Esta constatação pode dever-se ao facto das variáveis puderem ser dependentes, e ao controlá-las individualmente, pode fazer com que nenhuma delas defina apropriadamente a qualidade desse mesmo produto.

Quando existem múltiplos parâmetros do processo e estes são monitorizados simultaneamente, as cartas de controlo univariadas não devem ser usadas, pois as várias características da qualidade podem estar correlacionadas entre si, o que prejudica o desempenho das cartas em identificar uma alteração no processo.

Assim, quando se pretende controlar mais do que uma característica da qualidade de um produto, deve-se utilizar as cartas de controlo multivariadas, pois estas permitam reunir, em tempo real, todos os dados que por sua vez são utilizados para extrair informação importante, de forma a desenvolver esquemas eficientes para monitorizar o desempenho do processo.

Quando se detecta a presença de autocorrelação nos dados multivariados, deve-se ajustar o melhor modelo matemático ARIMA a estes dados e construir as cartas de controlo com os resíduos ou dos erros de previsão do modelo ajustado, tal como no estudo univariado. Deste

modo, para se construir as cartas multivariadas é necessário verificar, primeiramente, se as variáveis são autocorrelação, pois caso sejam, deve-se determinar os resíduos correspondentes as  $p$  variáveis e só depois aplicar as cartas de controlo multivariadas.

Foram várias as cartas de controlo propostas para monitorizar mais do que uma característica (variável) da qualidade de um produto ou processo, sendo a carta  $T^2$  de *Hotelling* (desenvolvida em 1947) uma das mais populares. Tal como no estudo univariado, há que distinguir duas fases distintas, na construção destas cartas, a Fase 1 e a Fase 2. A Fase 1 corresponde ao estudo retrospectivo do processo, no qual determinam-se os limites de controlo e verifica-se se o processo está sob controlo estatístico. A Fase 2 corresponde à monitorização do processo.

De seguida, apresentam-se as cartas multivariadas para o controlo da média das diversas variáveis.

### **Carta $T^2$**

A carta  $T^2$ , introduzida por *Hotelling* em 1947, é considerada como uma extensão multivariada da carta de controlo de *Shewhart* univariada baseada na monitorização das médias de amostras independentes. Para esta carta, recomenda-se um mínimo de 100 medições de cada característica da qualidade.

Para construir a carta  $T^2$ , deve-se distinguir as situações em que os dados são agrupados em amostras e em que os dados são observações individuais. Segundo *Lowry* e *Montgomery* (1995), o uso apropriado da estatística  $T^2$  ou  $\chi^2$  pode ser dividido em quatro categorias:

- 1) Fase 1 e  $n = 1$ , trabalha-se com observações individuais;
- 2) Fase 1 e  $n > 1$ , trabalha-se com subgrupos racionais;
- 3) Fase 2 e  $n = 1$ , trabalha-se com observações individuais;
- 4) Fase 2 e  $n > 1$ , trabalha-se com subgrupos racionais.

Nesta dissertação, apenas serão abordadas as situações em que os dados correspondem a observações individuais.

Assim, a estatística  $T^2$  para as  $p$  características da qualidade, cada uma delas com  $m$  observações individuais, no instante  $i$ , é dada pela seguinte expressão (Pereira e Requeijo, 2008):

$$(T^2)_i = (\mathbf{X}_i - \bar{\mathbf{X}})\mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}) \quad (2.21)$$

O vector  $\mathbf{X}_i$  constituído pelas observações das  $p$  características no instante  $i$ , é dado por

$$\mathbf{X}_i = (X_{1i}, X_{2i}, \dots, X_{pi})^T \quad (2.22)$$



no qual assume-se que  $X_1, X_2, \dots, X_p$  são independentes e identicamente distribuídas segundo  $N_p(\mu, \Sigma)$ .

O vector média  $\bar{X}$  é definido por

$$\bar{X} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_p)^T \quad (2.23)$$

onde  $\bar{X}_j$  (com  $j = 1, 2, \dots, p$ ) é a média das  $m$  observações, para a característica da qualidade  $j$ .

$$\bar{X}_j = \frac{1}{m} \sum_{k=1}^m X_{jk} \quad (2.24)$$

Por sua vez, a matriz de covariância  $S$  é dada por

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} & \dots & S_{1p} \\ S_{21} & S_{22} & S_{23} & \dots & S_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ S_{p1} & S_{p2} & S_{p3} & \dots & S_{pp} \end{bmatrix} \quad (2.25)$$

onde os elementos  $S_{jj}$  da matriz  $S$ , correspondem às variâncias amostrais de cada uma das  $p$  características e os elementos  $S_{jh}$  correspondem às covariância entre duas características distintas.

$$S_{jj} = S_j^2 = \frac{1}{m-1} \sum_{k=1}^m (X_{ji} - \bar{X}_j)^2 \quad (2.26)$$

$$S_{ji} = \frac{1}{m-1} \sum_{i=1}^m (X_{ji} - \bar{X}_j)(X_{jh} - \bar{X}_h) ; j \neq h \quad (2.27)$$

Segundo *Tracy et al.* (1992), os limites de controlo da carta  $T^2$ , para a Fase 1, devem ser baseados em função da distribuição Beta, sendo estes definidos pela seguinte expressão:

$$LSC_{T^2} = \frac{(m-1)^2}{m} \beta_{\alpha, (\frac{p}{2}, \frac{m-p-1}{2})} \quad (2.28)$$

$$LIC_{T^2} = 0$$

onde  $\beta_{\alpha, (p/2, (m-p-1)/2)}$  representa o percentil à direita, para uma probabilidade  $\alpha$ , da distribuição Beta com parâmetros  $p/2$  e  $(m-p-1)/2$ . No entanto, caso não se tenha acesso a *software* que permita calcular o valor da distribuição Beta, *Tracy et al* (1992) sugere a sua determinação em função da distribuição de *Fisher*:

$$\beta_{\alpha, p/2, (m-p-1)/2} = \frac{\frac{p}{m-p-1} F_{\alpha; p, (m-p-1)}}{1 + \frac{p}{m-p-1} F_{\alpha; p, (m-p-1)}} \quad (2.29)$$

onde  $F_{\alpha;p,(m-p-1)}$  corresponde ao percentil à direita para uma probabilidade  $\alpha$ .

Para a Fase 2, relativamente à carta  $T^2$ , a sua estatística é dada pela expressão 2.21 e os seus limites de controlo por:

$$LSC_{T^2} = \frac{p(m+1)(m-1)}{m(m-p)} F_{\alpha;p,m-p} \quad (2.30)$$

$$LIC_{T^2} = 0$$

Quando os parâmetros  $\mu$  e  $\Sigma$ , são estimados a partir de um grande número de amostras preliminares ( $m > 100$ ), o limite superior desta carta pode ser dado pela seguinte fórmula (Montgomery, 2005):

$$LSC_{T^2} = \chi^2_{\alpha,p} \quad (2.31)$$

Nestas cartas, tal como nas cartas univariadas, a detecção de uma causa especial de variação é feita quando um valor de  $(T^2)_i$  ou  $(X^2)_i$  excede os limites de controlo da respectiva carta.

Por outro lado, tal como no caso univariado, as cartas de controlo multivariadas do tipo *Shewhart* ( $T^2$  ou  $X^2$ ) apenas utilizam a informação das observações recentes, tornando-as insensíveis para detectar alterações pequenas e moderadas dos parâmetros das  $p$  características de um processo. Assim, para resolver este problema, foi necessário desenvolver as cartas CUSUM e EWMA para o caso multivariado. Estas cartas, tal como as cartas CUSUM e EMWA univariadas, dependem, também, do  $ARL_{EC}$  e do desvio que se pretende detectar. Uma vez que existem  $p$  características, o desvio é dado pelo parâmetro de não centralidade  $\delta^2(\mu)$  ou  $\delta(\mu)$  designado como “distância estatística”, e definido por (Pereira e Requeijo, 2008)

$$\delta^2(\mu) \Sigma^{-1}(\mu - \mu_0) \quad (2.32)$$

onde

$\mu_0$  – vector média inicial do processo;

$\delta^2(\mu)$  – parâmetro de não centralidade.

### **Carta CUSUM multivariada para o controlo da média**

A carta CUSUM multivariada (MCUSUM), é representada pela estatística  $Y^*$ . Esta estatística é definida, no instante  $t$ , por (Pereira e Requeijo, 2008):

$$Y_t^* = (\mathbf{C}_t \Sigma^{-1} \mathbf{C}_t)^{1/2} \quad (2.33)$$

onde

$$\mathbf{C}_t = (\mathbf{C}_{t-1} + \mathbf{X}_t - \boldsymbol{\mu}_0) \left(1 - \frac{k}{C_t^*}\right) \text{ se } C_t^* \leq k \quad (2.34)$$

$$\mathbf{C}_t = \mathbf{0} \text{ se } C_t^* > k \quad (2.35)$$

em que

$$C_t^* = [(\mathbf{C}_{t-1} + \mathbf{X}_t - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\mathbf{C}_{t-1} + \mathbf{X}_t - \boldsymbol{\mu}_0)]^{1/2} \quad (2.36)$$

$$\mathbf{C}_0 = \mathbf{0} \text{ e } k > 0 \quad (2.37)$$

Detecta-se a ocorrência de uma causa especial de variação sempre que um valor de  $Y_t^* > h$ , no qual  $h$  corresponde ao limite de controlo. Como a determinação do ARL, para esta carta, depende dos parâmetros de não centralidade, *Crosier* (1988) recomenda  $k = \delta(\mu)/2$ , onde  $\delta^2(\mu)$  é dado pela expressão (2.32).

*Crosier* (1988) conclui que a carta MCUSUM é mais sensível que a carta  $T^2$  e a carta  $\mathcal{X}^2$ , no que diz respeito à detecção de alterações pequenas e moderadas do vector média.

#### **Carta EWMA multivariada para o controlo da média**

A carta EWMA multivariada (MEWMA), é representada pela estatística  $T^2$  e definida, no instante  $t$ , por (Pereira e Requeijo, 2008)

$$T_t^2 = \mathbf{Z}_t^T \boldsymbol{\Sigma}_Z^{-1} \mathbf{Z}_t \quad (2.38)$$

onde

$$\mathbf{Z}_t = \mathbf{R}\mathbf{X}_t + (\mathbf{I} - \mathbf{R})\mathbf{Z}_{t-1}, \mathbf{Z}_0 = \mathbf{0} \quad (2.39)$$

$$\boldsymbol{\Sigma}_Z = \frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}] \boldsymbol{\Sigma} \quad (2.40)$$

Normalmente,  $\lambda_1 = \lambda_2 = \dots = \lambda_p$ . Quando isto acontece,  $\mathbf{Z}_t$  passa a ser definido por

$$\mathbf{Z}_t = \lambda \mathbf{X}_t + (1-\lambda)\mathbf{Z}_{t-1} \quad (2.41)$$

Nas expressões anteriores, tem-se:

$\boldsymbol{\Sigma}_Z$  – matriz covariância de  $\mathbf{Z}$ ;

$\mathbf{I}$  – matriz identidade;

$\mathbf{R} = \text{dig}(\lambda_1, \lambda_2, \dots, \lambda_p)$  com  $0 \leq \lambda_j \leq 1$ ;

$\lambda_j$  – constante de amortecimento para a variável  $j$  ( $j = 1, 2, \dots, p$ ).

Sempre que um valor de  $T_t^2 > H$ , detecta-se uma situação fora de controlo. O valor de  $H$  é determinado por simulação, de forma a encontrar um  $ARL_{EC}$  específico. O desempenho do  $ARL$ , depende apenas dos parâmetros de não centralidade, mas nos casos em que se utiliza constantes com pesos diferentes, o  $ARL$  passa a depender da direcção das alterações (Bersimis et al., 2006).

Alguns autores sugerem uma aproximação da expressão (2.40), dada por  $\Sigma_Z = (\lambda/(2 - \lambda))\Sigma$ . Contudo, o uso da expressão 2.40 conduz a uma resposta inicial rápida para a carta MEWMA (Bersimis et al., 2006).

### 2.2.3.1. Interpretação da Carta $T^2$

A identificação de uma causa especial nem sempre é fácil, pois tanto pode estar relacionada com uma única variável como a um subconjunto de variáveis. Este problema de detecção da(s) variável(eis) que esta(ão) associada(s) à existência de uma causa especial de variação foi investigada por vários autores. Alt, em 1985, propôs o uso de cartas univariadas das  $p$  variáveis com limites baseados nas desigualdades de Bonferroni. Jackson (1985) usou a decomposição de  $T^2$  ou  $(\chi^2)$  em componentes principais. Hauter e Tsui, em 1994, propuseram o método do ponto crítico,  $C_{R,\alpha}$ . Mason, Tracy e Young (1995, 1997) utilizaram a decomposição de  $T^2$  ou  $(\chi^2)$  em componentes ortogonais. Runger, Alt e Montgomery (1996) e Montgomery (2005) usaram a estatística  $d$ . Nedumaran e Pignatiello, em 1998, propuseram o uso de cartas univariadas com limites de diagnóstico.

Nesta dissertação apenas serão abordados dois dos métodos referidos anteriormente, o método da estatística  $d$  e o da decomposição em componentes ortogonais. De seguida, apresentam-se os dois métodos.

#### Método da Estatística $d$

Este método tem como base a determinação da variável/eis responsável/eis por uma situação de fora de controlo. Para tal, utiliza a estatística  $d$ , que é designada, para cada variável  $i$ , por (Pereira e Requeijo, 2008)

$$d_i = T^2 - T_{(i)}^2, \quad i = 1, 2, \dots, p \quad (2.42)$$

onde

$T^2$  – valor da estatística determinado com base nas  $p$  variáveis;

$T_{(i)}^2$  – valor da estatística calculada a partir de  $(p - 1)$  variáveis, tendo em conta que a variável em questão  $(i)$  não entra para o cálculo.

Uma variável contribui significativamente para a situação de fora de controlo se

$$d_i > \chi_{\alpha;1}^2 \quad (2.43)$$

### **Método de Decomposição em Componentes Ortogonais**

*Mason, Tracy e Young* (1995) sugerem decompor a estatística  $T^2$  em duas partes independentes (expressão 2.44), no qual cada uma delas reflecte a contribuição de uma variável individual. Esta decomposição ajuda os investigadores a detectar qual a/s variável/eis, com contribuição significativa, é/são a/as causadora/as do desvio existente.

$$T^2 = T_{p-1}^2 + T_{p,1,2,\dots,p-1}^2 \quad (2.44)$$

O primeiro termo da expressão 2.44 corresponde à estatística  $T^2$  das  $(p - 1)$  variáveis, no qual também pode ser separado em duas partes

$$T_{p-1}^2 = T_{p-2}^2 + T_{p-1,1,2,\dots,p-2}^2 \quad (2.45)$$

Este mesmo raciocínio pode ser feito para as estatísticas  $T_{p-k}^2$ , obtendo-se assim a equação equivalente à (2.44)

$$T^2 = T_1^2 + T_{2,1}^2 + T_{3,1,2}^2 + T_{4,1,2,3}^2 + \dots + T_{p,1,2,\dots,p-1}^2 = T_1^2 + \sum_{j=1}^{p-1} T_{j+1,1,2,\dots,j}^2 \quad (2.46)$$

onde

$T_1^2$  – estatística da primeira variável;

$\sum_{j=1}^{p-1} T_{j+1,1,2,\dots,j}^2$  – estatística da variável  $(j + 1)$  ajustada pelas estimativas da média e desvio padrão da distribuição condicionada da variável  $(j + 1)$ , dada a primeira variável  $j$ .

Existem  $p!$  combinações possíveis para decompor a estatística  $T^2$ . Por exemplo, caso se disponha de três variáveis ( $p = 3$ ), então existem seis possibilidades para decompor a estatística  $T^2$  (Tabela 2.2), no qual todas as combinações deverão ser analisadas (Pereira e Requeijo, 2008).

**Tabela 2.2** – Decomposição de  $T^2$  em componentes ortogonais para  $p=3$ 

	Ordenação das variáveis			Decomposição em Componentes Ortogonais
	1	2	3	
<b>Possibilidade 1</b>	$X_1$	$X_2$	$X_3$	$T_1^2 + T_{2,1}^2 + T_{3,1,2}^2$
<b>Possibilidade 2</b>	$X_1$	$X_3$	$X_2$	$T_1^2 + T_{3,1}^2 + T_{2,1,3}^2$
<b>Possibilidade 3</b>	$X_2$	$X_1$	$X_3$	$T_2^2 + T_{1,2}^2 + T_{3,2,1}^2$
<b>Possibilidade 4</b>	$X_2$	$X_3$	$X_1$	$T_2^2 + T_{3,2}^2 + T_{1,2,3}^2$
<b>Possibilidade 5</b>	$X_3$	$X_1$	$X_2$	$T_3^2 + T_{1,3}^2 + T_{2,3,1}^2$
<b>Possibilidade 6</b>	$X_3$	$X_2$	$X_1$	$T_3^2 + T_{2,3}^2 + T_{1,3,2}^2$

Os termos incondicionais ( $T_j^2$ ) e condicionais ( $T_{j+1,1,1,2,\dots,j}^2$ ) apresentados na tabela anterior podem ser dados pelas seguintes expressões

$$T_j^2 = \frac{(X_{ij} - \bar{X}_j)^2}{S_j^2} \quad (2.47)$$

$$T_{p,1,2,\dots,p-1}^2 = \frac{(X_{ip} - \bar{X}_{p,1,\dots,p-1})^2}{S_{p,1,\dots,p-1}^2} \quad \text{no qual } p = j + 1 \quad (2.48)$$

onde

$$\bar{X}_{p,1,\dots,p-1} = \bar{X}_p + \mathbf{b}_p^T (\mathbf{X}_i^{(p-1)} - \bar{\mathbf{X}}_i^{(p-1)}) \quad (2.49)$$

$$\mathbf{b}_p = \mathbf{S}_{XX}^{-1} \mathbf{s}_{Xp} \quad (2.50)$$

$$S_{p,1,\dots,p-1}^2 = S_X^2 - \mathbf{s}_{Xp}^T \mathbf{S}_{XX}^{-1} \mathbf{s}_{Xp} \quad (2.51)$$

Para se analisar uma situação de fora de controlo, utilizando este método, comparam-se os valores dos termos incondicionais e dos condicionais com os seus valores críticos. Assim, considera-se que um termo condicional contribui significativamente para a situação de fora de controlo se

$$T_{j+1,1,1,2,\dots,j}^2 > \frac{m-1}{m} F_{\alpha,1,m-1} \quad (2.52)$$

Já em relação ao termo incondicional, considera-se que este contribui significativamente para uma situação de fora de controlo se

$$T_j^2 > \frac{m-1}{m} F_{\alpha,1,m-1} \quad (2.53)$$

Este método permite apresentar um elevado número de combinações, o que pode fazer com que haja vários termos condicionais significativos, dificultando a sua interpretação. Assim, uma forma rápida de reduzir o número de termos a analisar, é a determinação da diferença entre o valor da estatística  $T^2$ , da observação em questão, e os termos incondicionais:

$$T^2 - T_1^2, T^2 - T_2^2, \dots, T^2 - T_p^2 \quad (2.54)$$

De seguida, compara-se cada diferença ( $T^2 - T_j^2$ ) com o valor crítico  $[(m-1)/m]F_{\alpha, m-1}$ . Assim, os termos condicionais ( $T_{j+1,1,2,\dots,j}^2$ ) não contribuem significativamente a uma situação de fora de controlo, logo não precisam de ser analisados, se

$$T^2 - T_j^2 < \frac{m-1}{m} F_{\alpha, m-1} \quad (2.55)$$

Por sua vez, os termos condicionais ( $T_{j+1,1,2,\dots,j}^2$ ) contribuem significativamente para a situação de fora de controlo, logo devem ser analisados, se

$$T^2 - T_j^2 > \frac{m-1}{m} F_{\alpha, m-1} \quad (2.56)$$

Relativamente aos termos incondicionais ( $T_j^2$ ) estes devem sempre ser comparados com o valor crítico  $[(m-1)/m]F_{\alpha, m-1}$ .

Por fim, se um termo incondicional apresentar um valor superior ao valor crítico significa que a situação de fora de controlo é devido a uma alteração no parâmetro da variável  $j$ . Já se um termo condicional apresentar um valor superior ao valor crítico, então a situação de fora de controlo é devido a uma alteração na correlação entre as variáveis representadas no termo em questão.

#### 2.2.4. Cartas de Controlo Multivariavas Baseadas em Métodos de Projecção

Quando se pretende controlar um número elevado de variáveis, as cartas multivariadas tradicionais ( $T^2$ , MCUSUM, MEWMA) tendem a perder a eficácia no que respeita à detecção de alterações no processo, uma vez que estas só são exequíveis quando o número de variáveis a monitorizar não é muito elevado (Matos, 2006). Assim, é necessário aplicar métodos que reduzam a dimensão do conjunto de variáveis. Os métodos mais comuns são os métodos de projecção, nomeadamente, a análise de componentes principais (PCA – *Principal Component Analysis*) e a regressão pelos Quadrados Mínimos Parciais (PLS – *Partial Least Square*). Na presente dissertação, e tendo em conta o que se pretende realizar na aplicação prática, apenas será desenvolvido o método de análise de componentes principais (PCA).

A análise de componentes principais (adiante, designada por PCA) é uma ferramenta poderosa na monitorização de processos complexos pois permite avaliar o desempenho de toda a unidade do processo (tanto pode controlar as características da qualidade de um produto, como também controlar vários sensores de temperatura, entre outros). Os resultados fundamentais no desenvolvimento da PCA, como um instrumento de monitorização de processo foram devido a *MacGregor e Kourti* (1995), *Kourti e MacGregor* (1995), com a contribuição importante de *Dunia e Quin*, em 1998, *Negiz e Cinar*, em 1997, *Raich e Cinar*, em 1997, entre outros.

Segundo *Ferrer* (2007), a PCA é uma técnica estatística multivariada usada para reduzir a dimensão do espaço de monitorização, projectando a informação das variáveis originais para subespaços dimensionais, pequenos, definidos por poucas variáveis independentes (os componentes principais).

Para aplicar a PCA, considera-se uma matriz de dados  $X$  composta por  $p$  variáveis aleatórias. De forma a obter-se resultados satisfatórios, deve-se proceder à normalização das variáveis, uma vez que estas são medidas em unidades diferentes. Assim, a cada coluna subtrai-se a média e divide-se pelo desvio padrão da respectiva variável, o que faz com que as variáveis passem a ter média zero e variância unitária (*Ku et al.*, 1995). Isto, normalmente melhora a interpretação do modelo uma vez que todas as variáveis possuem o mesmo valor de média. Por sua vez, esta matriz de dados é transformada em  $A$  componentes principais.

Um componente principal (*score*) é formado por uma combinação linear das variáveis aleatórias originais, sendo esta combinação obtida a partir da matriz de covariância ( $\Sigma$ ). Cada componente principal descreve a porção da variância total apresentada pelo conjunto das  $p$  variáveis aleatórias. Segundo *MacGregor e Kourti* (1995) e *Kourti* (2005), o primeiro componente principal, definido como uma combinação linear  $t_1 = p_1^T X$ , é determinado de forma a reter a maior variância possível, sujeita a  $p_1 = 1$ . O segundo componente principal ( $t_2 = p_2^T X$ ), terá a segunda variância mais alta, sujeita a  $p_2 = 1$ , mas submetido à condição de que não deve ser correlacionado com o primeiro. Os restantes componentes principais são definidos do mesmo modo, até esgotar toda a variância disponível no conjunto de variáveis. Os vectores  $p_i$  correspondem aos vectores próprios da matriz covariância ( $\Sigma$ ) de  $X$ , dada pela expressão (2.57) e os valores próprios  $\lambda_i$  correspondentes transmitem a variância de cada componente principal, obtida pela expressão (2.58).

$$\text{cov}(X) = \frac{X^T X}{m - 1} \quad (2.57)$$

$$\text{Var}(t_i) = \lambda_i \quad (2.58)$$

Na prática, raramente é necessário calcular os  $p$  vectores próprios, uma vez que a maior parte da variabilidade nos dados consegue ser capturada nos primeiros componentes principais, ou seja, 2 ou 3 dos componentes principais são suficientes para explicar grande parte da variância



aplicada nos produtos. Assim, uma vez retidos os primeiros  $A$  componentes principais, a matriz de dados  $X$  é decomposta numa soma de matrizes, definida por (Kourti, 2005)

$$X = TP^T + E = \sum_{i=1}^k t_i p_i^T + E \quad (2.59)$$

onde

$T$  – matriz dos componentes principais;

$P$  – matriz dos vectores próprios;

$E$  – matriz residual (representa a informação que não é explicada pelo modelo PCA);

A determinação dos componentes principais e dos respectivos vectores próprios pode ser obtida através de vários métodos. Segundo Kourti e MacGregor (1995) e Kourti (2005), o algoritmo NIPALS (*Nonlinear Iterative Pastial Least Square*) é o método ideal pois permite determinar os componentes principais de uma forma sequencial, quando o número de variáveis é elevado.

O número de componentes principais necessários para representar de forma eficiente o conjunto de dados pode ser obtido através de vários métodos, nomeadamente, a percentagem de variância explicada, o método de Kaiser, a validação cruzada e o gráfico de componentes *versus* os respectivos valores próprios. Segundo MacGregor e Kourti (1995), o método considerado mais adequado corresponde à validação cruzada.

Através dos componentes principais (*scores*) e dos resíduos associados a cada observação, são derivadas duas estatísticas complementares (ortogonais ou independentes), a estatística  $T_A^2$  de Hotelling e a estatística  $Q$  ou SPE (erro de previsão quadrático).

A estatística  $T_A^2$  representa uma medida total da variação do processo, ou seja, utiliza os componentes principais retidos para controlar a variabilidade das principais fontes do processo. Caso um valor desta estatística exceda os limites de controlo, significa que a observação correspondente apresenta valores extremos anormais em algumas (ou todas) das suas  $p$  variáveis originais. Matematicamente, a estatística  $T_A^2$  pode ser construída com base nos primeiros  $A$  componentes principais (Kourti, 2005):

$$T_A^2 = \sum_{i=1}^A \frac{t_i^2}{s_{t_i}^2} = \sum_{i=1}^A \frac{t_i^2}{\lambda_i} \quad (2.60)$$

onde

$t_i$  – valor do componente principal  $i$ ;

$\lambda_i$  – valor próprio do componente principal  $i$ .

O limite superior de controlo da estatística  $T_A^2$ , baseado nos primeiros  $A$  componentes principais, é obtido recorrendo à distribuição de *Fisher*, e representado pela seguinte expressão (Kourti, 2005):

$$LSC(T_A^2) = \frac{A(m^2 - 1)}{m(m - A)} F_{\alpha, (A, m-A)} \quad (2.61)$$

onde

$F_{\alpha, (A, m-A)}$  – representa o ponto crítico da distribuição de *Fisher*, considerando um nível de significância de  $\alpha$ .

Para esta carta, também se pode aproximar o limite superior de controlo à distribuição Qui-Quadrado (expressão 2.31).

A segunda carta de controlo, designada por  $Q$ , mede a quantidade de variação que não é capturada pelo modelo PCA (Kano et al., 2001). Pois, apesar da maior parte da variabilidade dos dados ser explicada pelos primeiros componentes principais (2 ou 3), os restantes componentes principais podem fornecer informações úteis uma vez que estes também capturam variâncias apesar de pequenas. Matematicamente, a estatística  $Q$  pode ser dada pela seguinte expressão (Kourti, 2005):

$$Q = \sum_{j=1}^k (X_{j,novo} - \hat{X}_{j,novo})^2 \quad (2.62)$$

onde

$X_{j,novo}$  – valor da observação da variável  $j$ ;

$\hat{X}_{j,novo}$  – estimativa da variável  $j$  resultante do modelo PCA

Assumindo-se que os resíduos seguem uma distribuição normal multivariada, vários procedimentos podem ser utilizados para determinar o limite superior de controlo da estatística  $Q$ . Como, por exemplo, Jackson e Mudholkar, em 1979, mostraram que o limite de controlo, pode ser dado por

$$LSC(Q) = \theta_1 \left[ \frac{z_{\alpha} \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \quad (2.63)$$

onde

$$\theta_i = \sum_{j=A+1}^m (\lambda_j)^i = TR(E^k), \quad \text{com } i = 1, 2, 3 \quad (2.64)$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_1^2} \quad (2.65)$$

Nas equações anteriores considera-se:

$z_\alpha$  – desvio normal correspondente ao percentil à direita para  $\alpha$  de probabilidade;

$\alpha$  – nível de significância;

$\lambda_j$  – valores próprios da matriz de covariância dos resíduos;

$\theta_i$  – valor obtido com base nos valores próprios da matriz de covariância dos resíduos.

Por sua vez, *Nomikos e MacGregor* (1995) usam uma aproximação baseada na distribuição Qui-quadrada ponderada:

$$Q_i \sim g\chi_h^2 \quad (2.66)$$

onde

$h$  – número de graus de liberdade;

$g$  – factor multiplicativo.

Estes dois investigadores sugeriram um modo simples e rápido para estimar os parâmetros  $g$  e  $h$ , que é baseado na correspondência de momentos entre a distribuição  $g\chi_h^2$  e a distribuição da amostra de  $Q$ . A média ( $\mu = gh$ ) e a variância ( $\sigma^2 = 2g^2h$ ) da distribuição Qui-quadrado ponderada são comparadas com a média ( $b$ ) e variância ( $v$ ) da amostra de  $Q$ . Por isso, os parâmetros  $g$  e  $h$  são estimados através das seguintes equações:

$$\hat{g} = \frac{v}{2b} \quad (2.67)$$

$$\hat{h} = \frac{2b^2}{v} \quad (2.68)$$

Assim, o limite superior de controlo da estatística  $Q$ , para um nível de significância  $\alpha$  é dado por

$$LSC(Q) = \frac{v}{2b} \chi_{\left(\frac{2b^2}{v}\right), \alpha}^2 \quad (2.69)$$

Para verificar se o processo está sob controlo estatístico, a carta  $Q$  deve ser a primeira a ser avaliada. Caso todos os pontos estejam dentro dos limites de controlo das duas cartas [ $T_{Ai}^2 < LSC(T_A^2)$  e  $Q_i \leq LSC(Q)$ ], o processo é considerado sob controlo.

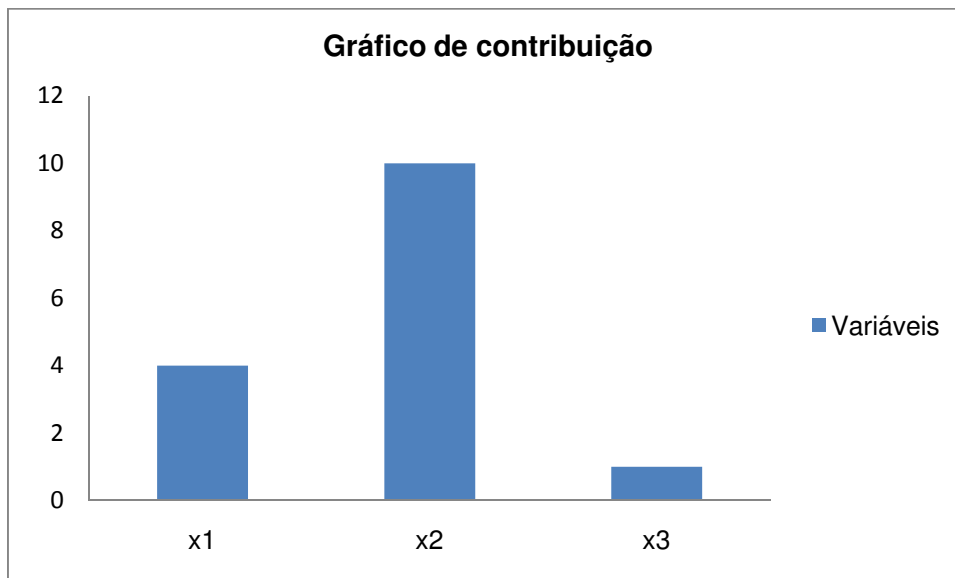
Segundo *Ostyn et al.* (2007), para que a construção destas duas cartas seja útil, na prática, devem considerar-se duas fases importantes. Numa primeira fase, o modelo PCA é construído usando um conjunto de dados de calibração e os valores limites das estatísticas  $T_A^2$  e  $Q$  são

estabelecidos usando equações (2.61) e (2.63). Numa segunda fase (validação), os novos dados são projectados no modelo PCA calibrado e as estatísticas  $T_A^2$  e  $Q$  são calculadas e comparadas com os valores-limites obtidos na primeira fase. Se as estatísticas ( $T_A^2$  ou  $Q$ ) traçadas estiveram acima destes limites superiores de controlo, significa que existe um desvio da observação, dando origem a um alarme (Montgomery, 2005).

#### 2.2.4.1. Interpretação da Carta PCA

Quando é detectado um ponto fora dos limites de controlo de uma das cartas, é necessário um diagnóstico isolado da variável original responsável pelo sinal fora de controlo. Um dos métodos mais utilizados para identificar quais as variáveis responsáveis pelo sinal fora de controlo, é o gráfico de contribuição. Este método é bastante poderoso para detectar situações fora de controlo pois utiliza a informação contida em todas as variáveis medidas simultaneamente.

Um gráfico de contribuição (Figura 2.3) indica quais as variáveis que contribuem numericamente para uma condição de fora de controlo. Estes são gráficos de barras no qual são desenhados os valores observados das variáveis, no período em que a carta sinaliza uma causa especial.



**Figura 2.3** – Exemplo de um gráfico de contribuição

Os gráficos de contribuição podem ser calculados para os pontos “anormais” nas duas cartas. Se a situação fora de controlo for detectada na carta  $Q$ , a contribuição de cada variável  $j$ , do conjunto de dados originais, é dada por (Kourti, 2005):

$$c_j^{(Q)} = (X_{j,novo} - \hat{X}_{j,novo})^2 \quad (2.70)$$

As variáveis com elevadas contribuições devem ser investigadas. Se a situação fora de controlo for detectada na carta  $T_A^2$ , é construído um gráfico de barras dos componentes principais normalizado  $[(t_i)^2/\lambda_i]$ , da observação em questão, e o componente principal que apresentar maior valor normalizado é seleccionado e investigado através do cálculo de contribuição da variável (Kourti, 2005). O gráfico de contribuição da variável indica como é que cada variável envolvida no cálculo deste componente principal, contribui para o sinal fora de controlo. Assim, a contribuição de cada variável do componente principal é dada por (Kourti, 2005):

$$C_j^{(t_i)} = p_{ij}(X_j - X_j^*) \quad (2.71)$$

onde  $p_{ij}$  representa o elemento da  $i$ -ésima linha e  $j$ -ésima coluna da matriz  $P$ .

Uma vez determinada a contribuição de cada variável, é construído o gráfico de contribuição. As variáveis, neste gráfico, que tiveram contribuições elevadas mas com o mesmo sinal que o componente principal devem ser investigadas. Por outro lado, as contribuições que tiverem sinal oposto ao do componente principal, apenas farão com que este seja menor.

Nos casos em que existem vários componentes principais com valores elevados, deve-se determinar a “contribuição média total” por variável para todos os componentes principais seleccionados, seguindo-se sempre as seguintes etapas (Kourti, 2005):

**1) Repetir para todos os  $k$  componentes principais elevados ( $k \leq A$ )**

- Determinar a contribuição da variável  $X_j$  para o componente principal seleccionado:

$$c_j^{(t_i)} = \frac{t_i}{\lambda_i} p_{ij}(X_j - X_j^*) \quad (2.72)$$

- Pôr a contribuição igual a zero caso seja negativa.

**2) Calcular a contribuição total da variável  $x_k$ .**

$$C_j^{(T_A^2)} = \sum_{i=1}^k (c_j^{(t_i)}) \quad (2.73)$$

**3) Investigar as variáveis com elevadas contribuições.**

O gráfico de contribuição pode não revelar a causa do evento especial, sendo deste modo necessário o conhecimento do processo por parte dos operadores de forma a poderem descobrir o que provocou as alterações nas variáveis identificadas.

### 2.2.4.2. Cartas de Controlo Multivariadas Baseadas na Análise de Componentes Principais Dinâmicas

Muitos são os processos que apresentam autocorrelação nos dados, principalmente na área da indústria química. Segundo *Tsung* (2000), para processos com autocorrelação significativa, a aplicação do método PCA convencional pode não ser eficiente pois como a base estatística desse método é perdida, devido aos dados violarem a suposição da independência do tempo, podem-se obter resultados insatisfatórios que, por sua vez, poderão gerar excessivos falsos alarmes, especialmente para perturbações de tamanhos pequenos. Além disso, quando há alterações no processo, a relação entre  $\varepsilon_t$  e  $X_t$  é sempre alterada. Deste modo, a aplicação directa do método PCA convencional nos dados, não irá revelar a relação exacta entre as variáveis, mas sim uma aproximação estática linear. Por isso, não é muito eficaz caracterizar a relação entre as variáveis segundo um modelo PCA estático.

Uma solução para este problema é a utilização do método PCA dinâmico (adiante designado por DPCA), proposto inicialmente por *Ku et al.* (1995). Estes investigadores estenderam o uso do método PCA na monitorização de processos para explicar a autocorrelação. Num sistema dinâmico, os valores actuais das variáveis vão depender dos valores passados por isso, é necessário identificar, no mínimo, a relação linear entre  $X(t)$  e  $X(t-1)$ . Este método é similar ao PCA original excepto a matriz dos dados ser composta por vectores replicados com atrasos de uma unidade de tempo por réplica. Assim, dependendo do número de atrasos ( $l$ ), a matriz de dados é expressa da seguinte forma:

$$X(l) = \begin{bmatrix} X_t^T & X_{t-1}^T & \cdots & X_{t-l}^T \\ X_{t-1}^T & X_{t-2}^T & \cdots & X_{t-l-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ X_{t+l-m}^T & X_{t+l-m-1}^T & \cdots & X_{t-m}^T \end{bmatrix} \quad (2.74)$$

onde  $X_t^T = [X_{1t}, X_{2t}, \dots, X_{pt}]^T$  é o vector de dimensão  $p$ , no instante  $t$ .

Segundo *Ku et al.* (1995), ao escolher correctamente o valor inteiro a atribuir ao número de atrasos ( $l$ ), os resíduos provenientes do modelo DPCA devem caracterizar-se como perfeitamente independentes, formando assim, uma base estatística estável para estudar um processo dinâmico.

Uma vez construída a matriz de dados, utilizam-se as mesmas expressões aplicadas no ponto 2.2.4 e ponto 2.2.4.1 para determinar as estatísticas  $T_A^2$  e  $Q$ , os respectivos limites de controlo bem como os gráficos de contribuições de forma a indicar as variáveis responsáveis pelas causas especiais.

Quando se aplica a PCA na matriz de dados dada pela expressão (2.74), identifica-se um modelo multivariado autorregressivo (AR) extraído directamente dos dados. Caso a matriz  $X(l)$

inclua as variáveis de entrada, então o modelo obtido será do tipo ARX – modelo autorregressivo com variável exógena (Ku *et al.*, 1995).

Segundo Russel *et al.* (2000), se na matriz  $X(l)$  são incluídos um número de atrasos ( $l$ ) suficientes, então a carta  $Q$  é estatisticamente independente de um instante de tempo para outro e o limite de controlo desta carta, dado pela expressão (2.63) é teoricamente justificado.

## 2.3. Síntese

Neste capítulo foi apresentado, numa primeira fase, uma breve revisão sobre o controlo estatístico do processo, no qual se expôs as cartas de controlo de Shewhart e as cartas de controlo especiais, nomeadamente a carta CUSUM e a EWMA. No que diz respeito às cartas de Shewhart, constatou-se que estas cartas deixam de ser úteis quando se pretende detectar alterações pequenas e moderadas dos parâmetros do processo, utilizando-se nestes casos as cartas CUSUM e EWMA.

Cada vez mais os processos industriais, nomeadamente processos químicos, tornam-se mais complexos, daí haver necessidade de automatizá-los e recolher os dados em tempo real. Tal facto potencia a presença de autocorrelação nos dados pois estes estão pouco espaçados no tempo. Assim, neste capítulo também se abordou o controlo estatístico do processo com dados autocorrelacionados. Foram apresentadas cartas de controlo univariadas e multivariadas concebidas para este tipo de processo.

Para o estudo univariado, foram consideradas duas abordagens de construção de cartas de controlo. A primeira abordagem não precisa de estabelecer um modelo matemático, no qual se referiram várias cartas, como por exemplo, a carta MCEWMA e a carta EWMAST. Já a segunda abordagem, necessita de um modelo matemático para determinar os valores previstos (ou resíduos), no qual se anunciaram algumas cartas, nomeadamente, a carta SCC e as cartas CUSUM e EWMA ambas de resíduos.

Já para o estudo multivariado, constata-se que cartas de controlo multivariadas tradicionais ( $T^2$ , MCUSUM, MEWMA) são de extrema importância nos processos indústrias quando se pretende controlar mais do que uma variável, podendo estas ser aplicadas aos dados autocorrelacionados. Mas em contrapartida, quando o número de variáveis a controlar é elevado, estas cartas tendem a perder a eficácia no que respeita à detecção de alterações no processo, sendo necessária a aplicação dos métodos de projecção. A carta de controlo baseada na análise de componentes principais dinâmica (DPCA) demonstrou-se bastante útil, quando os dados são autocorrelacionados, pois permite eliminar a autocorrelação, evitando assim a modelação do processo.

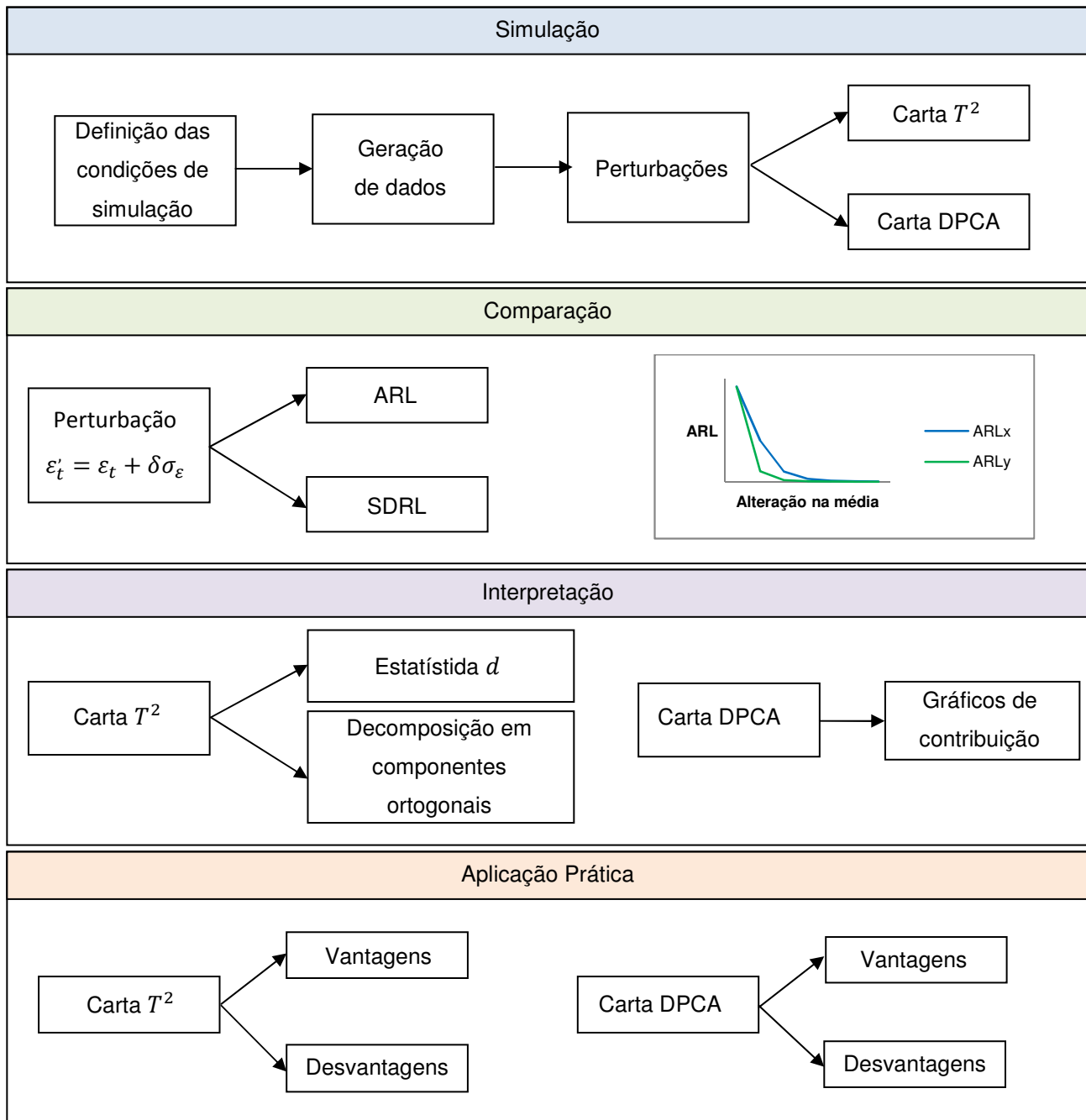
### 3. Metodologia

Neste capítulo, sugere-se uma metodologia para a realização do estudo comparativo de duas cartas de controlo multivariadas com dados autocorrelacionados. Para uma melhor compreensão esta metodologia será apresentada recorrendo a fluxogramas e esquemas tendo por base a informação disponível no Capítulo 2. Este estudo será aplicado somente à análise de cartas de controlo multivariadas de observações individuais e dados autocorrelacionados.



### 3.1. Apresentação da Metodologia Proposta

Uma vez que o principal objectivo da presente dissertação consiste em desenvolver uma metodologia que permita estabelecer uma comparação entre o desempenho de cartas de controlo multivariadas, apresenta-se na Figura 3.1, um esquema que permite indicar as várias etapas que irão caracterizar a metodologia a adoptar.

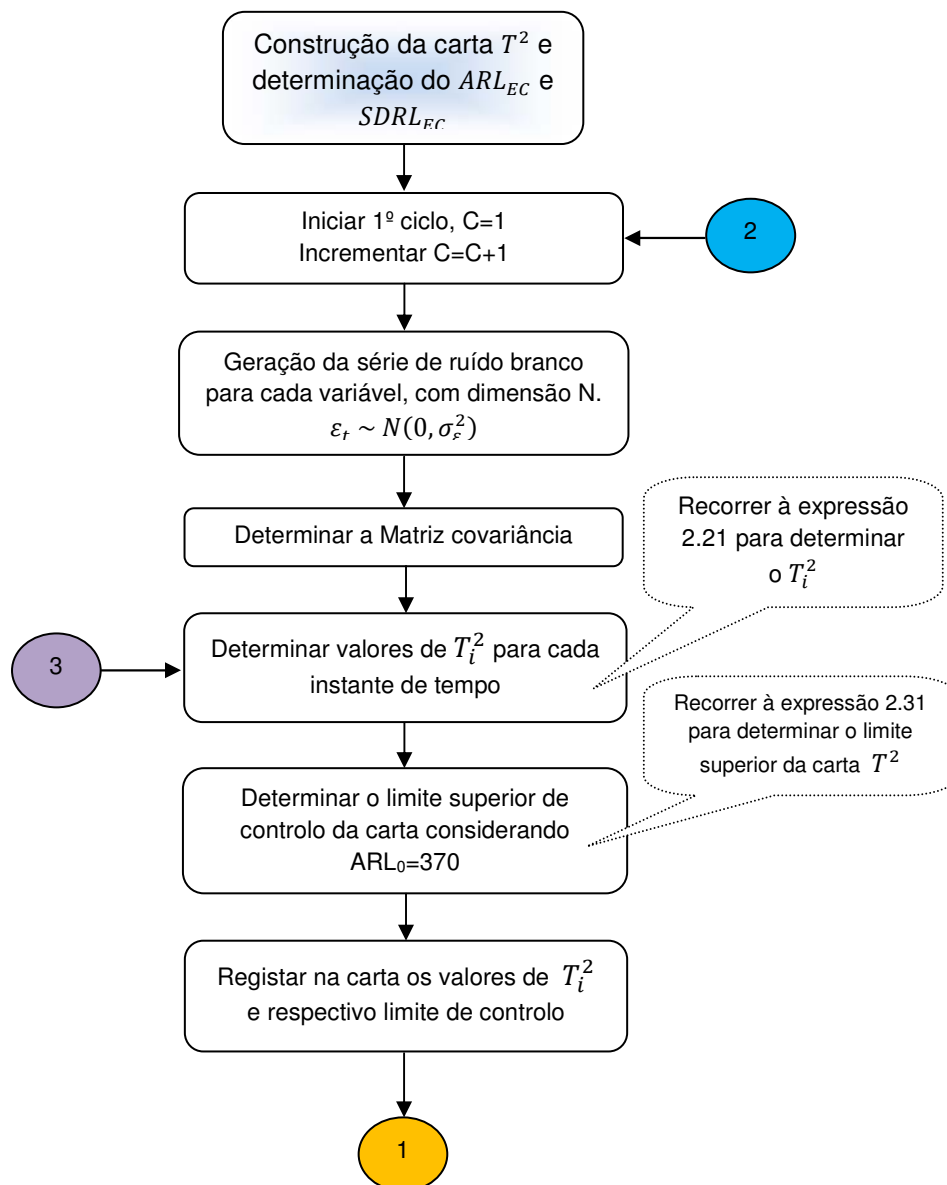


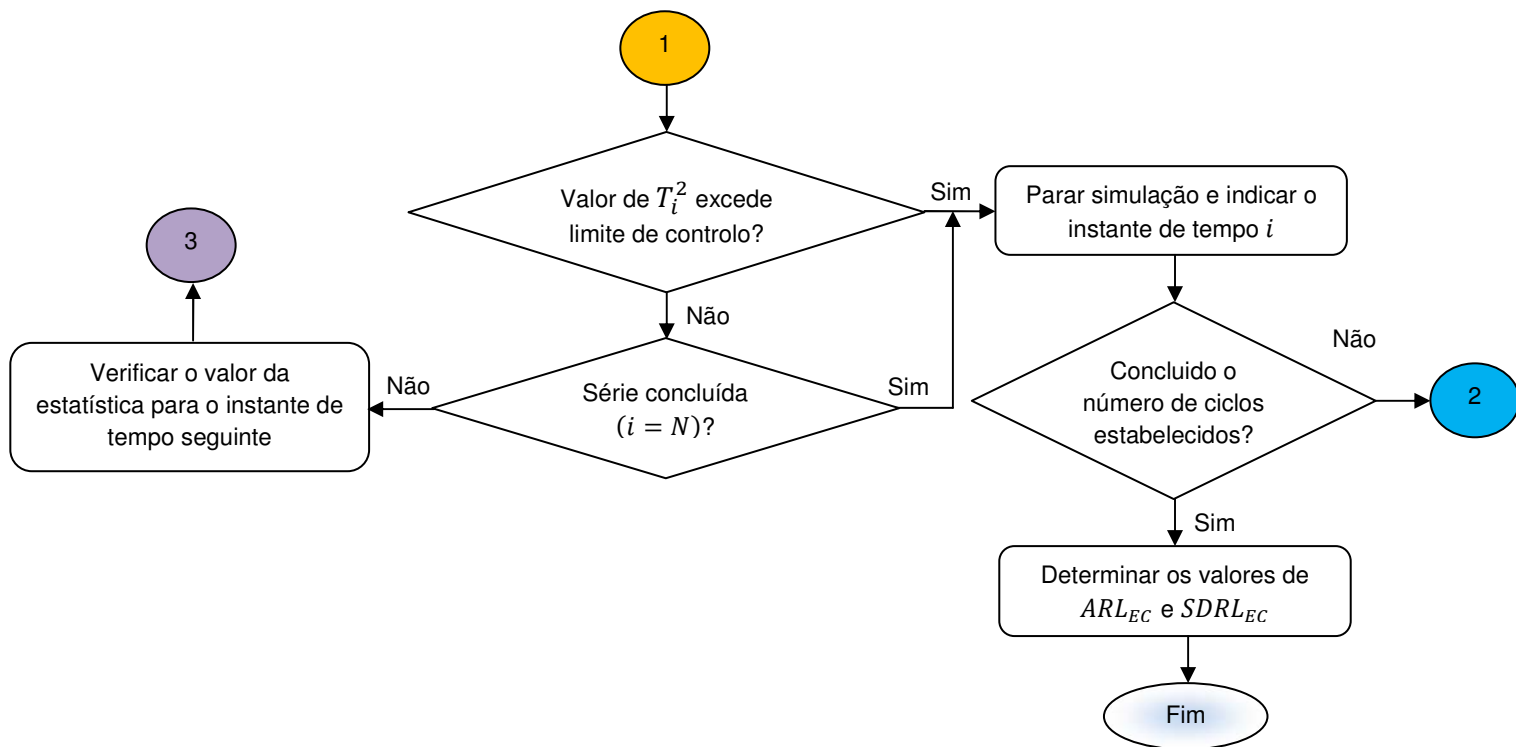
**Figura 3.1** – Fases que compõem a metodologia a adoptar

### 3.2. Construção do Programa de Simulação

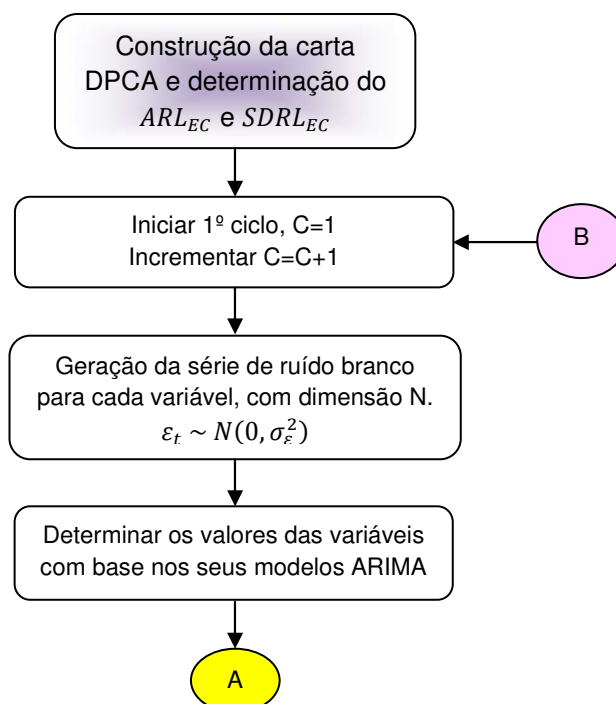
Os modelos de simulação serão construídos num programa desenvolvido sobre a plataforma do *software* MATLAB. Um aspecto a ter em consideração, quando se realizam simulações, é a definição do número de ciclos a efectuar bem como o número de observações de cada simulação, de forma a garantir a qualidade dos resultados que se pretende obter.

Os esquemas que permitem auxiliar na definição dos modelos de simulação estão representados nas figuras seguintes.





**Figura 3.2** – Construção da carta  $T^2$  e determinação do  $ARL_{EC}$  e  $SDRL_{EC}$



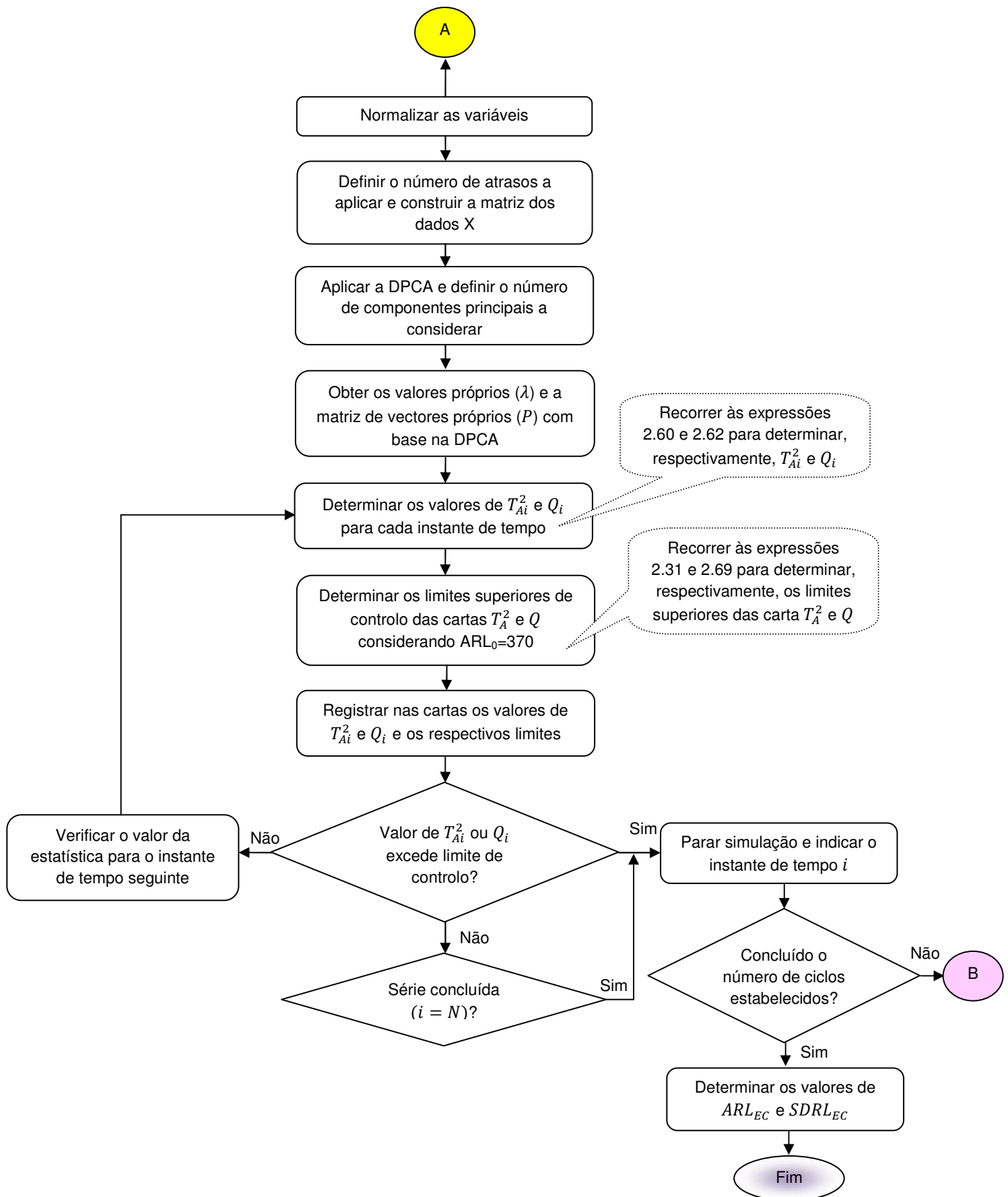


Figura 3.3 – Construção da carta DPCA e determinação do  $ARL_{EC}$  e  $SDRL_{EC}$

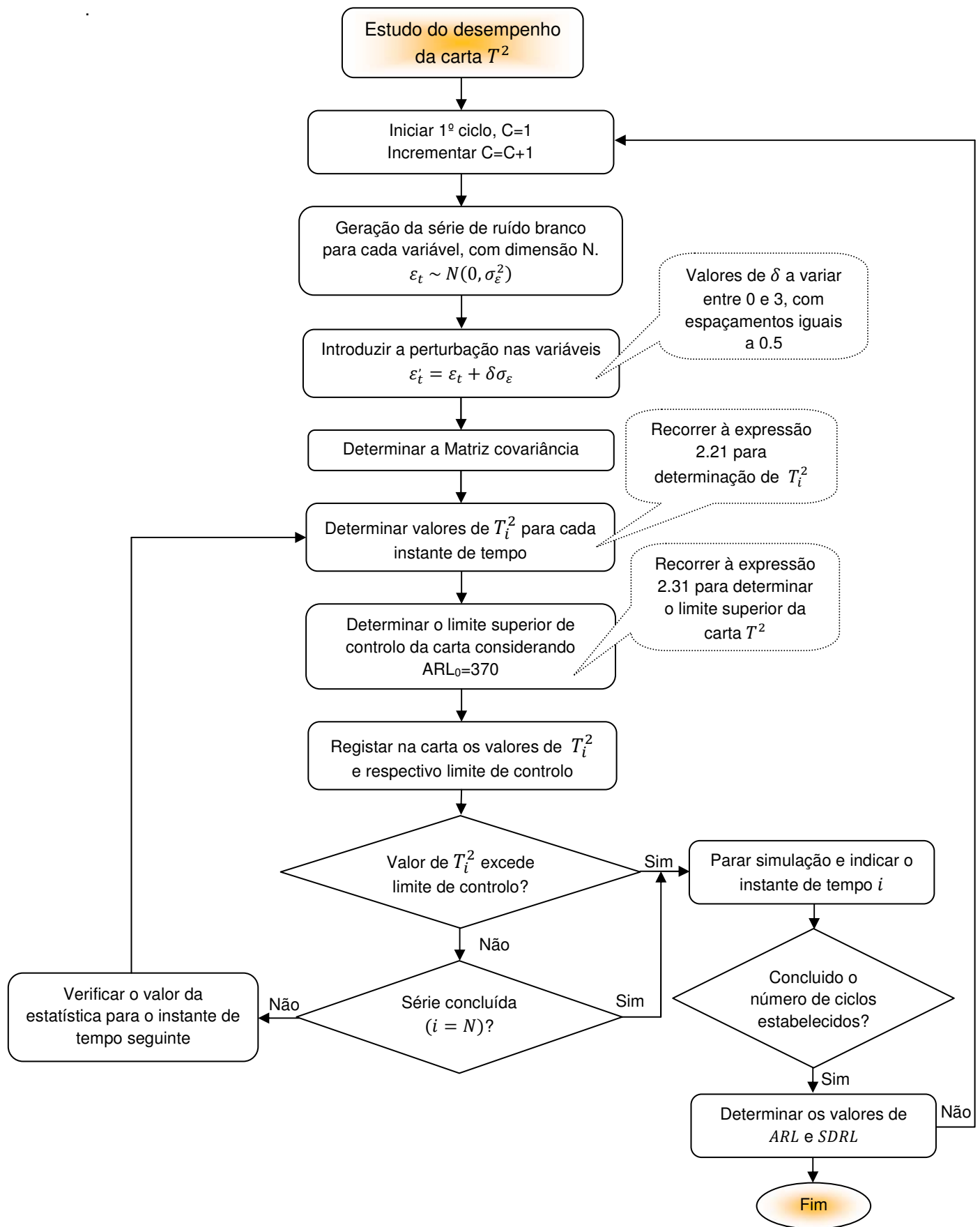
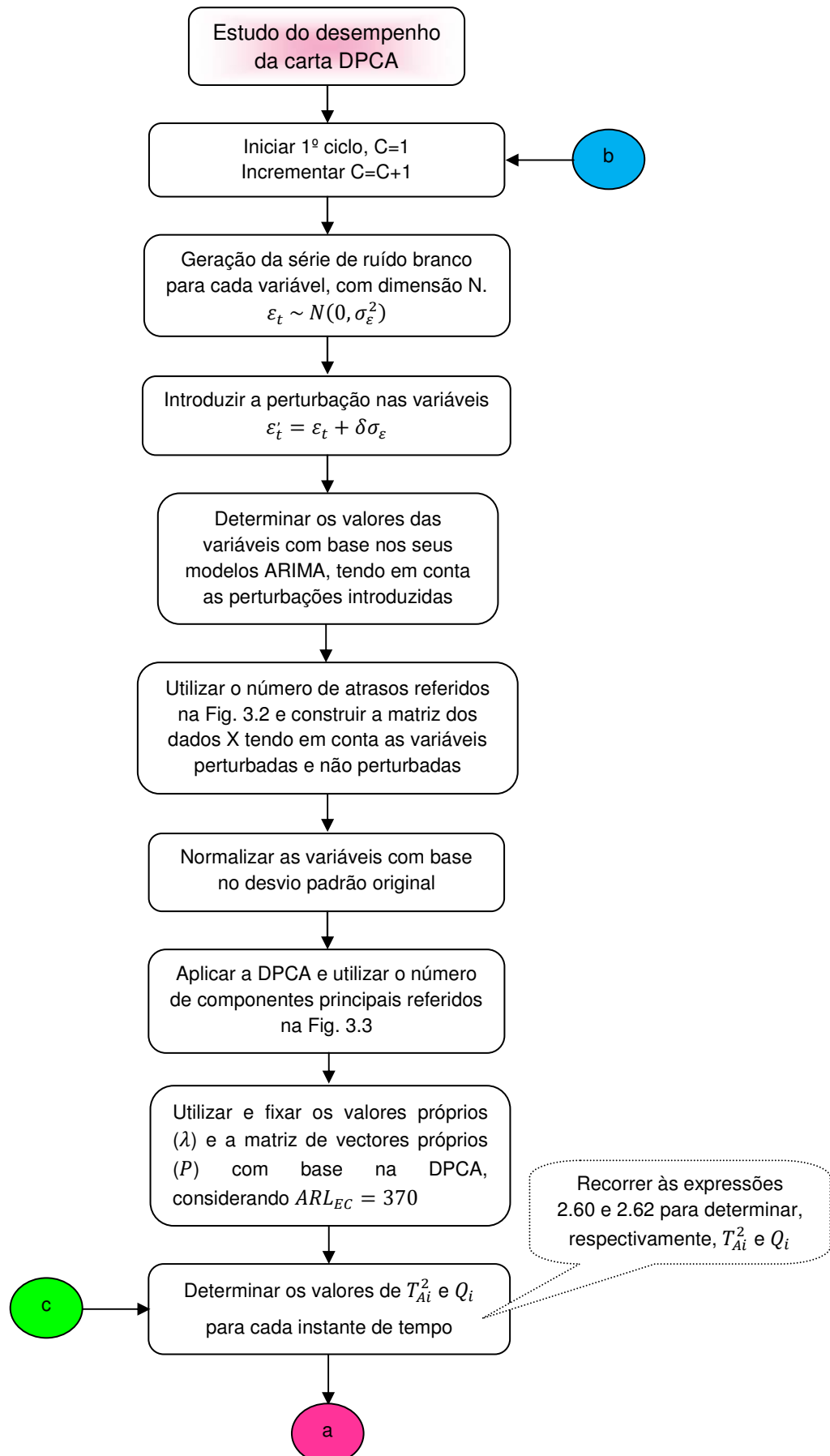
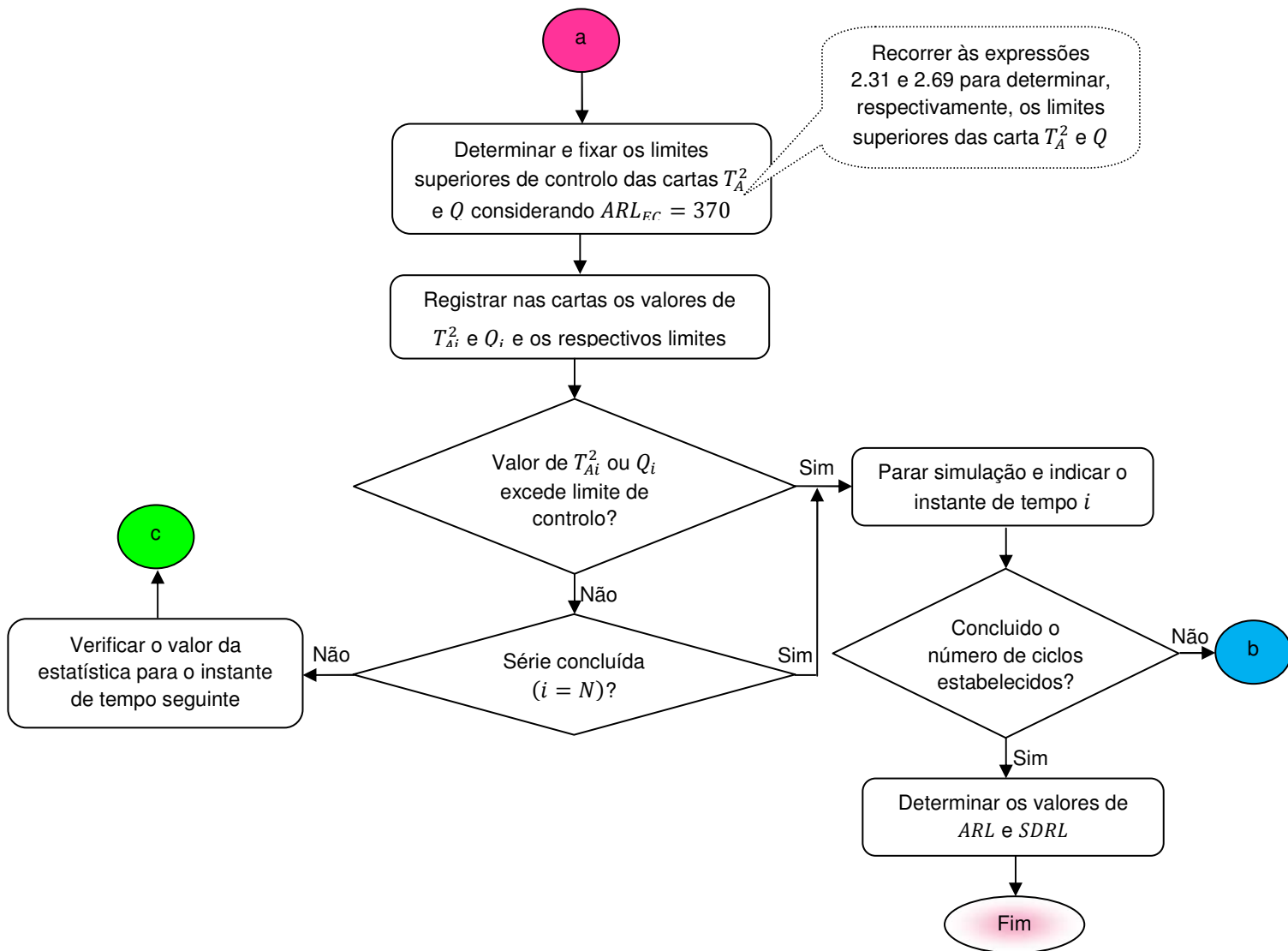


Figura 3.4 – Estudo do desempenho da carta  $T^2$  a alterações no parâmetro médio





**Figura 3.5** – Estudo do desempenho da carta DPCA a alterações no parâmetro médio

### 3.3. Comparação do Desempenho das Cartas de Controlo

Para a realização da análise comparativa do desempenho das cartas de controlo multivariadas, vai ser considerado o valor do ARL, assim como o valor do respectivo SDRL (desvio padrão). Na maioria dos estudos, a análise dos valores de SDRL é ignorada, mas a sua interpretação, juntamente com o valor de ARL, assume um papel importante. O valor de SDRL permite auxiliar na decisão da melhor carta pois se uma carta, apesar de apresentar um valor de ARL mais baixo, tiver um valor de SDRL maior em comparação às outras cartas, significa que esta carta apresenta um comportamento instável. Este estudo comparativo servirá para indicar qual a melhor carta, quando o processo é sujeito a um tipo de perturbação numa ou mais variáveis em simultâneo.

### 3.4. Interpretação das Cartas de Controlo Multivariadas

Quando existe um ponto fora dos limites de controlo, é necessário indicar quais as variáveis responsáveis por esta situação de fora de controlo. Na presente dissertação serão utilizados três métodos para a identificação das variáveis responsáveis por uma situação de fora de controlo, nomeadamente, a estatística  $d$ , a decomposição em componentes ortogonais e os gráficos de contribuição. De seguida, apresenta-se um esquema que servirá de base para a aplicação dos três métodos.

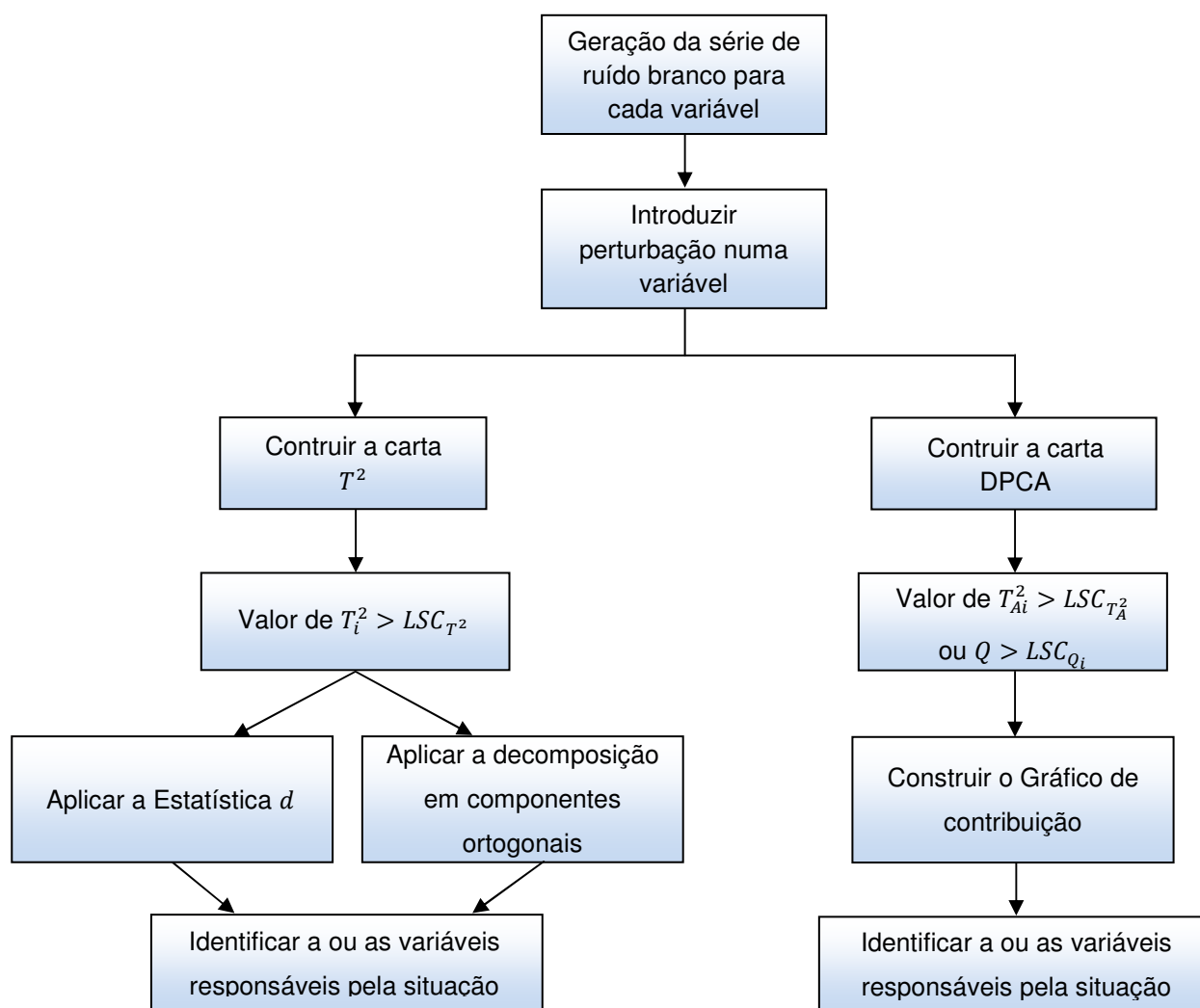


Figura 3.6 – Interpretação de cartas multivariadas



### **3.5. Aplicação Prática**

A aplicação prática corresponde à última fase da metodologia proposta. Esta fase tem como principal objectivo identificar as vantagens e desvantagens das cartas de controlo na óptica do utilizador que irá aplicar a carta.

## 4. Desenvolvimento Prático

Este capítulo tem como principal objectivo analisar o desempenho de duas cartas de controlo multivariadas com dados autocorrelacionados. No primeiro ponto são apresentados os modelos matemáticos de três variáveis, obtidos a partir de dados de um processo real. No segundo ponto, realiza-se uma análise de sensibilidade das duas cartas, nomeadamente a carta  $T^2$  e a carta DPCA, através da medida de desempenho mais utilizada, ARL. No terceiro, apresenta-se o desenvolvimento de três metodologias de identificação de causas especiais para as duas cartas multivariadas. No quarto ponto apresentam-se as vantagens e desvantagens das duas cartas. Por fim é elaborada uma síntese dos desenvolvimentos mais relevantes do presente capítulo.

### 4.1. Identificação das Variáveis do Processo

A abordagem proposta na presente dissertação foi a escolha de três variáveis AR(1) com parâmetros ( $\theta$ ) e erro quadrático médio ( $MS_e$ ) diferentes. Considerou-se que três variáveis seriam um número acessível para obter resultados satisfatórios, dada a complexidade do estudo que se pretende desenvolver neste trabalho. Os modelos matemáticos que representam estas três variáveis foram retirados de dados reais autocorrelacionados, obtidos de um processo produtivo químico. De seguida apresentam-se os modelos matemáticos das três variáveis.

Variável 1 ( $X1$ ) :  $X_t = 0,614X_{t-1} + \varepsilon_t$  com  $MS_e = 7,99$

Variável 2 ( $X2$ ) :  $X_t = 0,647X_{t-1} + \varepsilon_t$  com  $MS_e = 3,10$

Variável 3 ( $X3$ ) :  $X_t = 0,451X_{t-1} + \varepsilon_t$  com  $MS_e = 10,61$

A determinação de cada série de ruído branco ( $\varepsilon_t$ ) é feita com o auxílio do *software* STATISTICA, no qual se gera um número aleatório de observações, tendo em conta o desvio padrão, a média e o número de observações que se pretende obter para cada variável. Na Figura 4.1 está representada a variável  $X1$ , de forma a visualizar-se o andamento do processo. Esta variável foi obtida através de simulação, em que se utilizou 1000 observações para a série gerada. Analisando a Figura 4.1, verifica-se que esta apresenta um comportamento estacionário, característico de um modelo autorregressivo de 1ª ordem.

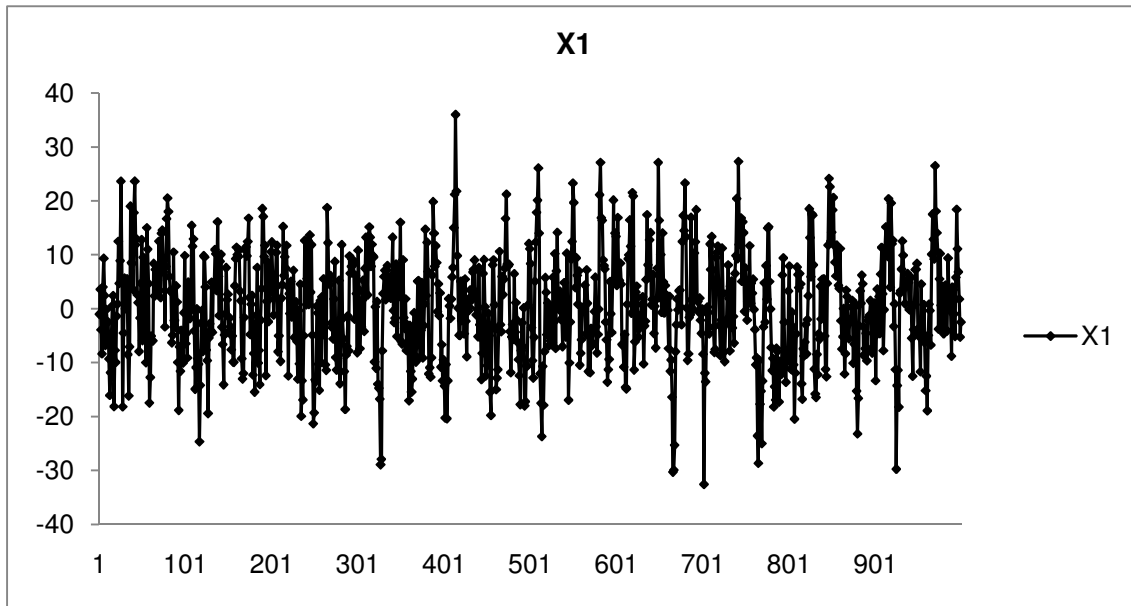


Figura 4.1 – Comportamento da carta X1

## 4.2. Análise de Sensibilidade das Cartas Através do ARL

Uma vez que se pretende analisar e comparar o desempenho de duas cartas de controlo multivariadas com dados autocorrelacionados ( $T^2$  e DPCA), será realizado um estudo de análise de sensibilidade, considerando alterações na média do processo, tendo como medida de desempenho o ARL, assim como o respectivo SDRL. Para que a realização deste estudo seja possível, é necessário que as duas cartas de controlo sejam construídas sob a mesma plataforma, isto é, que possuam um valor de  $ARL_{EC}$  igual, visto que a base de construção de cada uma das cartas é diferente. Deste modo, apresenta-se de seguida a determinação do  $ARL_{EC}$  e respectivo  $SDRL_{EC}$  para as duas cartas.

### 4.2.1. Determinação do $ARL_{EC}$ e do $SDRL_{EC}$

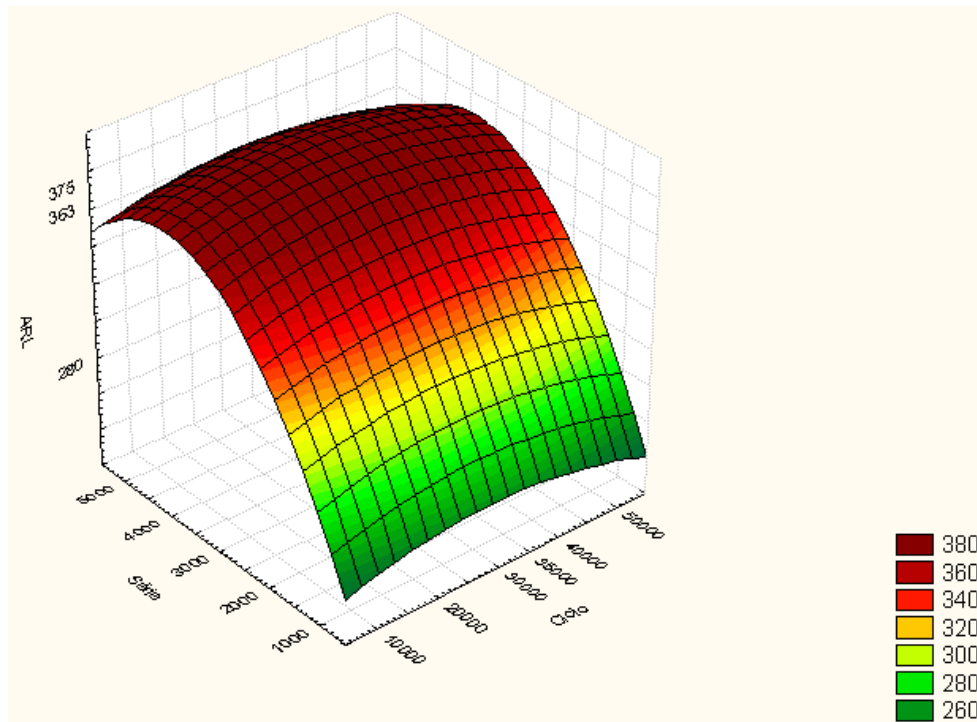
Para se determinar o  $ARL_{EC}$  e o  $SDRL_{EC}$  através de modelos de simulação, é necessário, primeiramente, definir o número de ciclos a efectuar bem como o número de observações em cada ciclo, de forma a garantir a qualidade dos resultados que se pretende obter. Relativamente ao número de observações de cada simulação ( $N$ ), deve-se ter em consideração o valor de  $ARL_{EC}$  e  $SDRL_{EC}$ . Já em relação ao número total de ciclos, é necessário ter igualmente em consideração o tempo necessário à realização de cada ciclo (depende igualmente de  $N$ ).

Deste modo, utilizou-se o modelo de simulação desenvolvido para a carta  $T^2$  (Figura 3.2) para se determinar os valores de  $ARL_{EC}$  e  $SDRL_{EC}$ . à medida que se variava o número de ciclos e a dimensão da série. Como se pretende minimizar o número de ciclos de forma a reduzir o tempo de realização das simulações, também se teve em conta o tempo de ciclo de cada simulação. Os resultados obtidos das simulações estão apresentados na seguinte tabela.

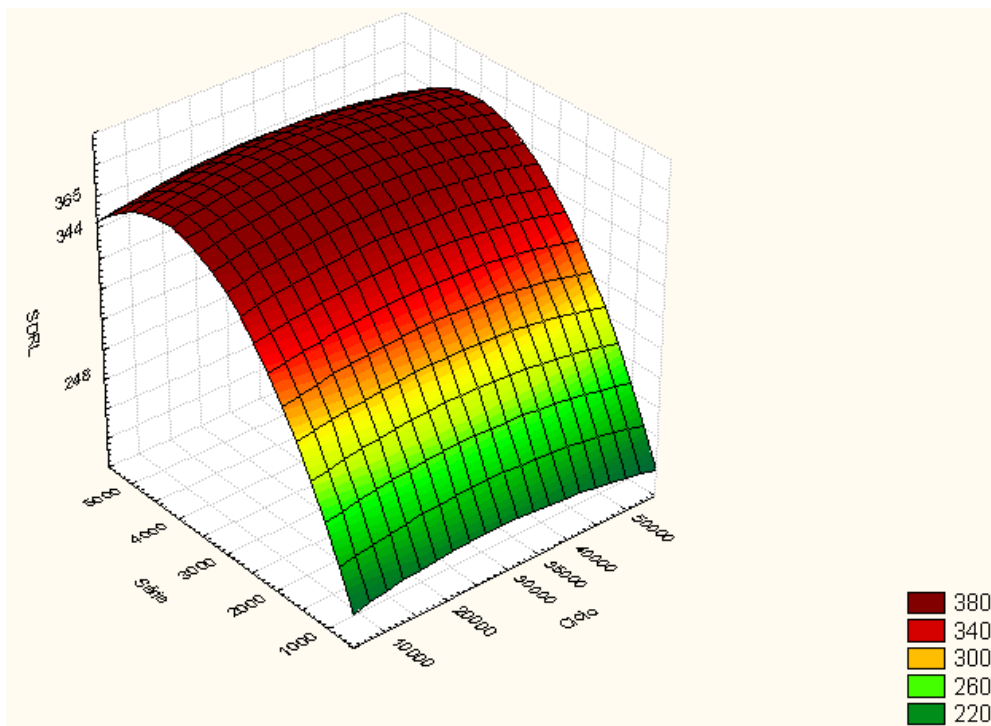
**Tabela 4.1** – Valores de  $ARL_{EC}$  e  $SDRL_{EC}$  para diferentes números de ciclos e dimensão da série

Ciclo (C)	Série (N)	ARL	SDRL	$T_{ciclo}$
10000	1000	274,0	243,0	36
20000	1000	282,0	246,0	72
30000	1000	278,0	244,0	105
40000	1000	278,0	255,0	165
50000	1000	279,0	246,0	181
20000	2000	359,0	341,0	83
30000	2000	357,0	339,0	121
20000	3000	367,0	364,0	93
30000	3000	366,0	360,0	144
20000	4000	372,0	366,0	104
30000	4000	368,0	365,0	154
40000	4000	369,0	367,0	211
20000	5000	370,0	368,0	106
30000	5000	370,0	372,0	173
35000	5000	370,0	370,0	200
40000	5000	370,0	370,0	217

Analizando a Tabela 4.1, e tendo em conta que o número de ciclos e a dimensão da série a utilizar deverá coincidir com a estabilização dos valores de ARL e SDRL, conclui-se que devem ser utilizados em cada simulação, 35.000 ciclos e cada ciclo deverá ser constituído por 5000 observações. Para se reforçar esta conclusão, efectuaram-se dois gráficos de superfície 3D (Figura 4.2 e 4.3), com recurso ao *software* STATISTICA. No primeiro variou-se o número de ciclos, dimensão da série e os valores de ARL e no segundo variou-se o número de ciclos, dimensão da série e os valores de SDRL.



**Figura 4.2** – Gráfico de Superfície 3D em que se varia o valor de ARL, do número de ciclos e da dimensão da série



**Figura 4.3** – Gráfico de Superfície 3D em que se varia o valor de SDRL, do número de ciclos e da dimensão da série

Numa primeira fase, pensou-se comparar a carta  $T^2$  e a  $T_A^2$  (baseada na DPCA). Para tal, determinou-se vários valores de  $ARL_{EC}$ , recorrendo a modelos de simulação no MATLAB, de forma a verificar se através destas duas cartas se obteria o mesmo valor de  $ARL_{EC}$ . Como esta fase foi de exploração, as simulações foram realizadas com séries iguais a 5000 mas com ciclos mais pequenos, iguais a 10 000, de forma a obter-se resultados mais rápidos. Apresenta-se de seguida uma tabela resumo com os resultados destas simulações.

**Tabela 4.2** – Determinação dos valores de  $ARL_{EC}$  e respectivo  $SDRL_{EC}$  para a carta  $T^2$  e  $T_A^2$  alterando o nível de significância

			Carta $T^2$		Carta $T_A^2$	
$\alpha$	$1/\alpha$	LSC	$ARL_{EC}$	$SDRL_{EC}$	$ARL_{EC}$	$SDRL_{EC}$
0,001	1000	16,26	972	906	1131	1044
0,0025	400	14,32	406	409	523	522
0,0027	370	14,16	369	372	476	471
0,005	200	12,84	202	199	276	269
0,0075	133,3	11,97	137	133	185	184
0,01	100	11,34	104	101	141	140
0,025	40	9,35	41	40	61	60

Analisando a Tabela 4.2, verifica-se que os valores de  $ARL_{EC}$  da carta  $T^2$  correspondem aos valores que se pretendiam obter. O mesmo não acontece com os valores de  $ARL_{EC}$  da carta  $T_A^2$ . Tal facto indica que esta carta não funciona correctamente sem a sua carta complementar, ou seja, a carta  $Q$ . Assim, o estudo do  $ARL_{EC}$  e respectivo  $SDRL_{EC}$  para a carta DPCA será feito com base na carta  $T_A^2$  e na carta  $Q$ , por forma a poder comparar-se os dois métodos, isto é, a carta  $T^2$  e a carta DPCA.

Os dois modelos de simulação, para determinação do  $ARL_{EC}$  e respectivo  $SDRL_{EC}$ , foram construídos de forma a obter-se um valor de  $ARL_{EC}$  igual a 370 pois considerou-se que este valor corresponderia a um bom compromisso entre a probabilidade de ocorrência de um falso alarme e a probabilidade em detectar a ocorrência de uma causa especial de variação. Os esquemas que serviram de base para a realização destes dois modelos encontram-se no Capítulo 3 (Figura 3.2 e Figura 3.3). As simulações, nos dois modelos, foram efectuadas com 35 000 ciclos ( $C$ ), sendo cada ciclo constituído por um total de 5000 observações ( $N$ ), tendo-se obtido os seguintes valores de  $ARL_{EC}$  e de  $SDRL_{EC}$  :

**Tabela 4.3** – Valores de  $ARL_{EC}$  e de  $SDRL_{EC}$  para as duas cartas

	<b>Carta <math>T^2</math></b>	<b>Carta DPCA</b>
<b>Nível de significância (<math>\alpha</math>)</b>	$\alpha_{T^2} = 0,0027$	$\alpha_{T_A^2} = 0,0027 \quad \alpha_Q = 0,0016$
<b>Limite superior de controlo (LSC)</b>	$LSC_{T^2} = 14,156$	$LSC_{T_A^2} = 14,156 \quad LSC_Q = 6,820$
<b><math>ARL_{EC}</math></b>	370,290	370,430
<b><math>SDRL_{EC}</math></b>	364,960	368,520

O nível de significância para a carta  $Q$  foi determinado recorrendo a iterações, de forma a obter-se um valor de  $ARL_{EC} = 370$ .

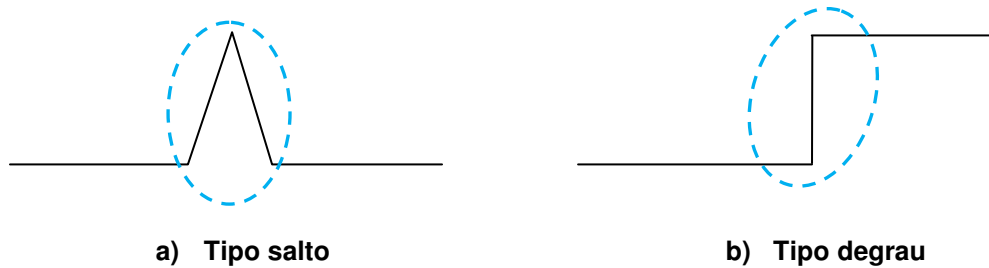
Como se pode verificar, quando se utiliza a carta  $T_A^2$  e  $Q$  consegue-se obter o valor de  $ARL_{EC}$  desejado, ou seja,  $ARL_{EC} = 370$ . Um valor de  $ARL_{EC} = 370$  significa que de 370 em 370 pontos existe, em média, um falso alarme, o que é desejável pois assim não haverá um número elevado de intervenções desnecessárias no processo. Uma vez que ambas as cartas em estudo possuem o mesmo valor de  $ARL_{EC}$ , pode-se proceder à análise comparativa do desempenho.

#### 4.2.2. Introdução de Perturbações na Média do Processo

Neste ponto, pretende-se introduzir uma perturbação na média do processo, de forma a verificar como é que cada uma das cartas reage, tendo como medida de desempenho o valor de  $ARL$  e respectivo desvio padrão ( $SDRL$ ). Uma das preocupações a ter em consideração é o tipo de perturbação que se deve utilizar e como introduzi-la nas duas cartas.

Relativamente aos tipos de perturbação que podem afectar um processo, podem-se considerar, como sendo os mais comuns, o tipo salto e o tipo degrau, representados na Figura 4.4. Uma perturbação do tipo salto, pode surgir com um gesto errado do operador, uma causa pontual que afectou num único instante ou uma peça que entrou no sistema e não devia, o que faz com que o processo recupere logo, dificultando a sua detecção. Já uma perturbação do tipo degrau, pode surgir com a avaria de um motor, de um sensor ou uma alteração na composição da matéria-prima, por exemplo, o que faz com que esta permaneça no processo, facilitando assim a sua detecção. Na presente dissertação, utiliza-se a perturbação do tipo degrau, uma vez se que se pretende saber a facilidade em detectar a perturbação introduzida, ponendo

esta ser detectada no próprio instante ou nos instantes seguintes. Já em relação a forma como se introduz a perturbação, chegou-se à conclusão que podiam ocorrer perturbações, em tempo real, tanto numa variável como em duas ou mesmo nas três variáveis (caso acontecesse um problema externo que pudesse alterar duas das variáveis ou mesmo as três). Assim, as perturbações serão introduzidas para estes três cenários.



**Figura 4.4** – Tipos de Perturbação

A alteração considerada no processo vai reflectir-se ao nível do parâmetro da média pela adição de um valor de  $\delta\sigma_\varepsilon$  a cada um dos valores gerados da série de ruído branco ( $\varepsilon_t$ ). O parâmetro  $\delta$  vai variar entre 0 e 3, com incrementos constantes de 0,5 unidades. Uma vez que se dispõe de três variáveis, a perturbação nos dois modelos de simulação será introduzida para sete situações:

- ✓ Variável 1
- ✓ Variável 2
- ✓ Variável 3
- }
- 1º cenário
- ✓ Variável 1 e 2
- ✓ Variável 1 e 3
- ✓ Variável 2 e 3
- }
- 2º cenário
- ✓ Variáveis 1, 2 e 3
- }
- 3º cenário

Para este estudo, mantêve-se a mesma condição para as duas cartas, isto é, um valor de  $ARL_{EC}$  igual a 370  $[\alpha_{T^2} = 0,0027; \alpha_{T_A^2} = 0,0027 \text{ e } \alpha_Q = 0,0016]$ .

#### 4.2.2.1. Desempenho da Carta $T^2$ a Alterações na Média do Processo

O modelo de simulação desenvolvido para a carta  $T^2$  teve por base o esquema que se apresenta no Capítulo 3 (Figura 3.2). As simulações foram efectuadas com 35 000 ciclos ( $C$ ), sendo cada ciclo constituído por um total de 5000 observações ( $N$ ). Relativamente à introdução de perturbações neste modelo, não houve qualquer problema, apresentando-se na Tabela 4.4,

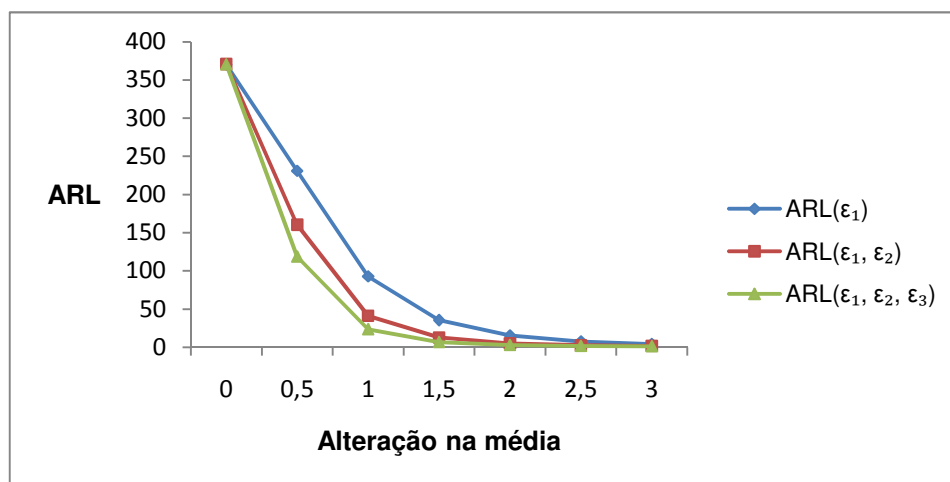


os resultados obtidos de ARL e respectivos SDRL, para três dos casos anteriormente enunciados. Os restantes encontram-se no ANEXO I (Tabela I.1 e I.2).

**Tabela 4.4** – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo considerando o estudo da carta  $T^2$

$\delta$	Carta $T^2$					
	Perturbação no $\varepsilon_1$		Perturbação no $\varepsilon_1$ e $\varepsilon_2$		Perturbação no $\varepsilon_1$ , $\varepsilon_2$ e $\varepsilon_3$	
	ARL	SDRL	ARL	SDRL	ARL	SDRL
0	370,23	370,00	370,87	367,37	370,43	364,86
0,5	230,99	230,68	160,42	159,76	118,56	117,87
1	92,69	92,41	41,14	40,66	23,42	22,94
1,5	35,55	34,99	12,74	12,16	6,74	6,28
2	15,46	14,89	5,14	4,62	2,84	2,29
2,5	7,64	7,12	2,69	2,14	1,68	1,07
3	4,30	3,78	1,75	1,15	1,25	0,56

Para uma melhor visualização do comportamento do valor do ARL quando é/são perturbada (s) uma, duas ou três variáveis, elaborou-se a seguinte figura:



**Figura 4.5** – Valores de ARL, para diferentes alterações no parâmetro médio do processo considerando o estudo da carta  $T^2$

Como base na interpretação gráfica da Figura 4.5. e analisando as Tabelas 4.4, I.1 e I.2, verifica-se que quando se introduz uma perturbação numa variável (1, ou 2 ou 3), os valores de ARL e respectivo SDRL em nada diferem, ou seja, os valores diminuem à medida que se aumenta a magnitude da perturbação, na mesma proporção. Já quando se introduz uma perturbação em duas variáveis (1 e 2 ou 1 e 3 ou 2 e 3), os valores de ARL e respectivos SDRL diminuem em relação aos valores obtidos quando se introduz a perturbação numa única variável. Por fim, quando se introduz uma perturbação nas três variáveis (1, 2 e 3) constata-se que os valores de ARL e respectivo SDRL também reduzem em relação aos valores obtidos quando a perturbação é introduzida em duas variáveis.

#### 4.2.2.2. Desempenho da Carta DPCA a Alterações na Média do Processo

O modelo de simulação que se desenvolveu para a carta DPCA, teve por base o esquema apresentado no Capítulo 3 (Figura 3.3) no qual se adicionou, apenas, um valor de  $\delta\sigma_\varepsilon$  a cada um dos valores gerados da série de ruído branco ( $\varepsilon_t$ ) e fixou-se o limite da carta. As simulações foram efectuadas com 35 000 ciclos ( $C$ ), sendo cada ciclo constituído por um total de 5000 observações ( $N$ ). Os resultados obtidos demonstraram que este modelo não funcionava correctamente pois quando se introduziu as perturbações, reparou-se que à medida que se aumentava a magnitude da perturbação ( $\delta$ ), o ARL aumentava, em vez de diminuir, como se pode verificar no exemplo apresentado na Tabela 4.5.

**Tabela 4.5** – Valores de ARL e SDRL para diferentes alterações no parâmetro médio do processo quando se introduz uma perturbação na variável 1, considerando o estudo da carta DPCA

Carta DPCA		
Perturbação no $\varepsilon_1$		
$\delta$	ARL	SDRL
0	370,43	368,52
0,5	388,76	388,16
1	429,24	431,68
1,5	495,98	503,88
2	540,86	563,39
2,5	633,86	678,77
3	672,54	743,34

Assim, chegou-se à conclusão que o valor de ARL aumentava à medida que se aumentava a magnitude da perturbação uma vez que a estrutura do modelo estava a ser alterada, isto é, estava-se a calcular novamente os valores próprios e os vectores próprios influenciando assim a matriz dos resíduos (a matriz de resíduos é a ferramenta principal para detectar o ARL) e por outro lado, a normalização da matriz dos dados estava a ser feita com o novo desvio padrão e não com o desvio padrão original.

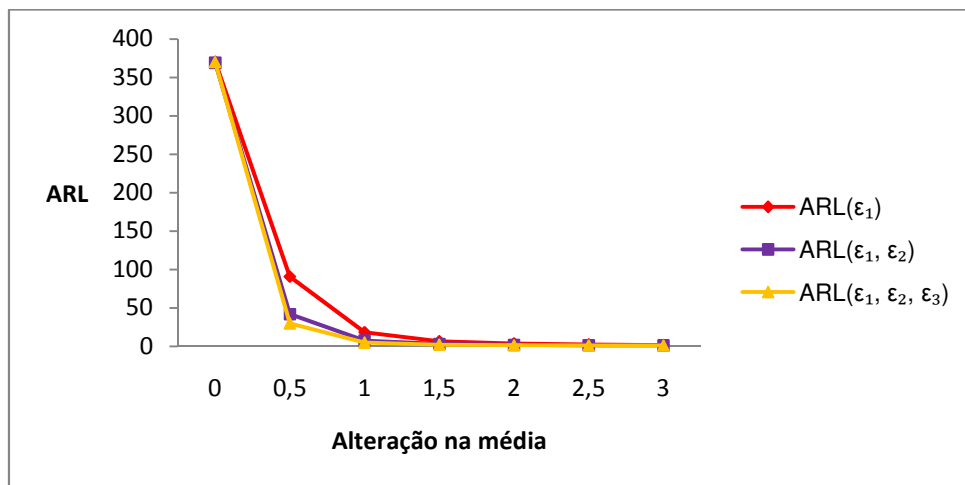
Para resolver este problema, fizeram-se várias simulações e verificou-se que o valor de ARL estava a depender da última coluna da matriz de vectores próprios, ou seja, se o terceiro valor desta coluna fosse pequeno, o valor do ARL também era, se o valor aumentasse, o ARL também aumentava. Perante estas conclusões, constatou-se que se devia fixar os valores próprios e vectores próprios. Assim, realizaram-se várias simulações de forma a determinar os valores próprios e vectores próprios que conduziam a um  $ARL_{EC} = 370$ .

Uma vez obtidos esses valores próprios e vectores próprios, construiu-se um novo modelo de simulação, que teve por base o esquema apresentado no Capítulo 3 (Figura 3.5). As simulações foram efectuadas com 35 000 ciclos ( $C$ ), sendo cada ciclo constituído por um total de 5000 observações ( $N$ ). Na Tabela 4.6 apresentam-se os resultados obtidos de ARL e respectivos SDRL, para três dos casos anteriormente enunciados. Os restantes casos encontram-se no ANEXO I (Tabela I.3 e I.4).

**Tabela 4.6** – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo, considerando o estudo da carta DPCA

$\delta$	Carta DPCA					
	Perturbação no $\varepsilon_1$		Perturbação no $\varepsilon_1$ e $\varepsilon_2$		Perturbação no $\varepsilon_1, \varepsilon_2$ e $\varepsilon_3$	
	ARL	SDRL	ARL	SDRL	ARL	SDRL
0	370,43	368,52	369,00	365,31	370,10	365,40
0,5	90,66	87,48	41,85	39,08	29,9	27,32
1	18,21	15,97	7,33	5,23	4,94	3,31
1,5	6,49	4,62	3,04	1,61	2,12	1,05
2	3,42	1,96	1,83	0,79	1,3	0,51
2,5	2,21	1,09	1,29	0,48	1,03	0,20
3	1,61	0,71	1,06	0,24	1,00	0,04

Para uma melhor visualização do comportamento do valor do ARL quando é/são perturbada (s) uma, duas ou três variáveis, elaborou-se a seguinte figura:



**Figura 4.6** – Valores de ARL, para diferentes alterações no parâmetro médio do processo considerando o estudo da carta DPCA

Analisando as Tabelas 4.6, I.3 e I.4 verifica-se que os valores de ARL e respectivos SDRL em nada diferem quando se introduz a perturbação na variável 1 ou na variável 2 ou na variável 3. O mesmo acontece quando se introduz perturbações em duas variáveis (1 e 2 ou 1 e 3 ou 2 e 3) e em 3 variáveis (1, 2 e 3). Por outro lado, à medida que se aumenta o número de variáveis perturbadas, os valores de ARL e respectivos SDRL diminuem, tal como acontece na carta  $T^2$ .

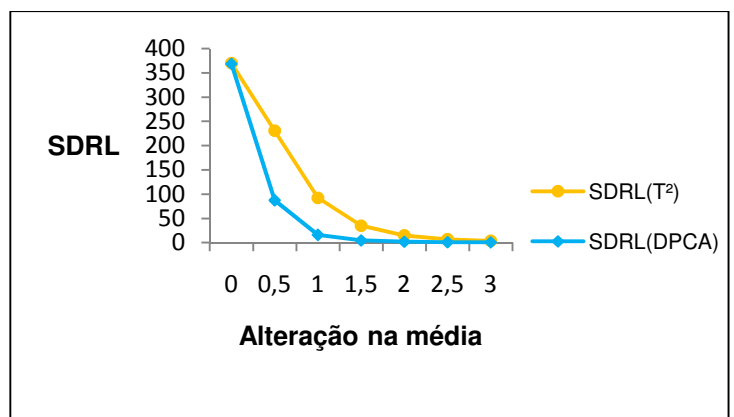
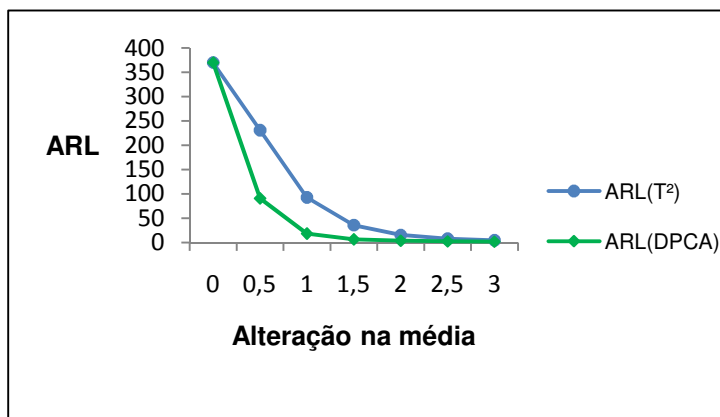
#### 4.2.2.3. Comparação do Desempenho das Duas Cartas

Para uma melhor comparação dos valores de ARL e SDRL, apresentam-se de seguida, as tabelas resumo para os três cenários de introdução de perturbações. Escolheu-se ao acaso os exemplos para cada cenário, uma vez que os valores de ARL e SDRL em nada diferem.

**1º cenário:** Introdução de perturbações numa variável**Tabela 4.7** – Valores de ARL e SDRL para diferentes alterações na média da variável 1

$\delta$	Perturbação no $\varepsilon_1$			
	Carta $T^2$		Carta DPCA	
	ARL	SDRL	ARL	SDRL
0	370,23	370,00	370,43	368,52
0,5	230,99	230,68	90,66	87,48
1	92,69	92,41	18,21	15,97
1,5	35,55	34,99	6,49	4,62
2	15,46	14,89	3,42	1,96
2,5	7,64	7,12	2,21	1,09
3	4,30	3,78	1,61	0,71

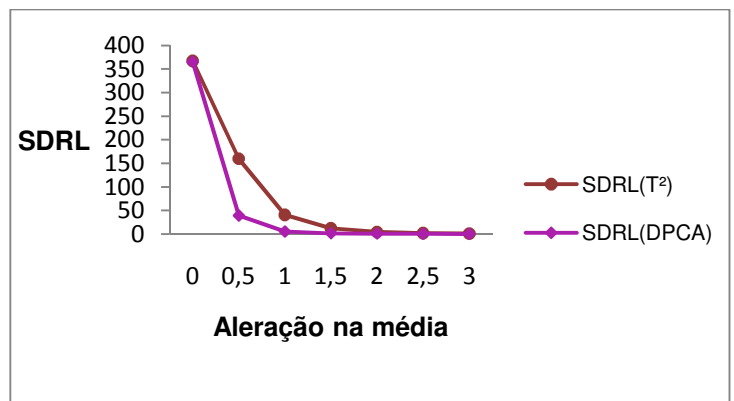
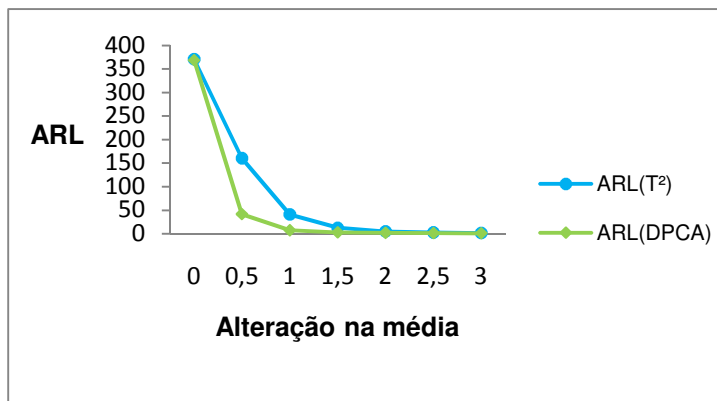
Para uma melhor visualização dos valores de ARL, assim como de SDRL, apresenta-se na Figura 4.7 o andamento de ambas as curvas.

**Figura 4.7** – Valores de ARL e SDRL para diferentes alterações na média da variável 1

**2º cenário:** Introdução de perturbações em duas variáveis**Tabela 4.8** – Valores de ARL e SDRL para diferentes alterações na média da variável 1 e 2

$\delta$	Perturbação no $\varepsilon_1$ e $\varepsilon_2$			
	Carta $T^2$		Carta DPCA	
	ARL	SDRL	ARL	SDRL
0	370,87	367,37	369,00	365,31
0,5	160,42	159,76	41,85	39,08
1	41,14	40,66	7,33	5,23
1,5	12,74	12,16	3,04	1,61
2	5,14	4,62	1,83	0,79
2,5	2,69	2,14	1,29	0,48
3	1,75	1,15	1,06	0,24

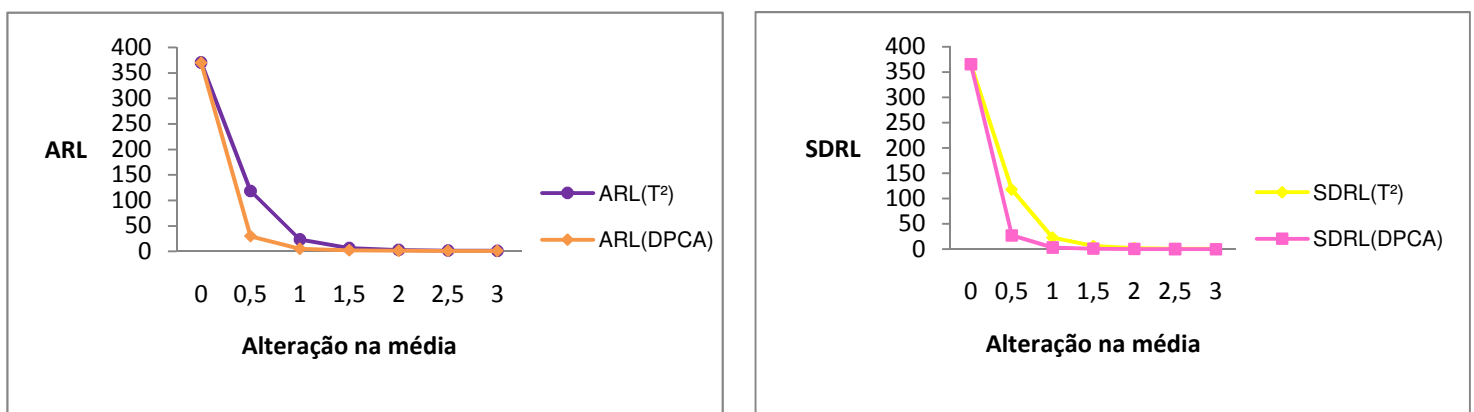
Para uma melhor visualização dos valores de ARL, assim como de SDRL, apresenta-se na Figura 4.8 o andamento de ambas as curvas.

**Figura 4.8** – Valores de ARL e SDRL para diferentes alterações na média da variável 1 e 2

**3º cenário:** Introdução de perturbações nas três variáveis**Tabela 4.9** – Valores de ARL e SDRL para diferentes alterações na média das três variáveis

$\delta$	Perturbação no $\varepsilon_1$ e $\varepsilon_2$ e $\varepsilon_3$			
	Carta $T^2$		Carta DPCA	
	ARL	SDRL	ARL	SDRL
0	370,43	364,86	370,10	365,40
0,5	118,56	117,87	29,90	27,32
1	23,42	22,94	4,94	3,31
1,5	6,74	6,28	2,12	1,05
2	2,84	2,29	1,3	0,51
2,5	1,68	1,07	1,03	0,20
3	1,25	0,56	1,00	0,04

Para uma melhor visualização dos valores de ARL, assim como de SDRL, apresenta-se na Figura 4.9 o andamento de ambas as curvas.

**Figura 4.9** – Valores de ARL e SDRL para diferentes alterações na média das três variáveis

Após uma análise conjunta das Tabelas e Figuras apresentadas para cada cenário, conclui-se que a carta DPCA é a mais eficaz pois apresenta o melhor valor de ARL, independentemente da perturbação ser introduzida numa variável, como em duas ou em três, permitindo assim detectar com maior rapidez alterações na média do processo, mesmo quando estas são pequenas (0,5), médias (1 a 1,5) ou elevadas (2 a 3). Por outro lado, esta carta também é a que apresenta um valor de SDRL mais baixo para cada cenário.

Uma vez que se concluiu que a carta DPCA é mais sensível para detectar alterações na média do processo, considerou-se que seria importante verificar se esta carta também detecta mais falsos alarmes que a carta  $T^2$ . Este estudo foi feito a partir da mesma geração de resíduos, com recurso ao *software* STATISTICA. A perturbação foi introduzida no mesmo instante ( $t = 25$ ) e na mesma variável, com uma magnitude igual a 1,5. A carta  $T^2$  foi construída com base nos resíduos obtidos, já a carta DPCA foi construída a partir da série original. De seguida apresentam-se as duas cartas.

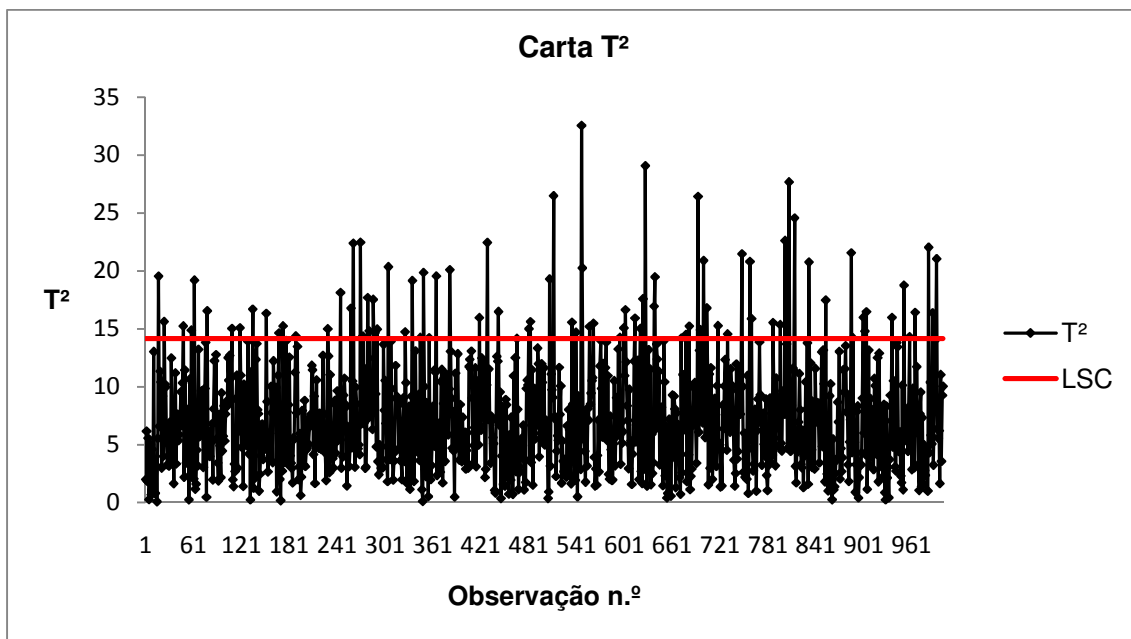
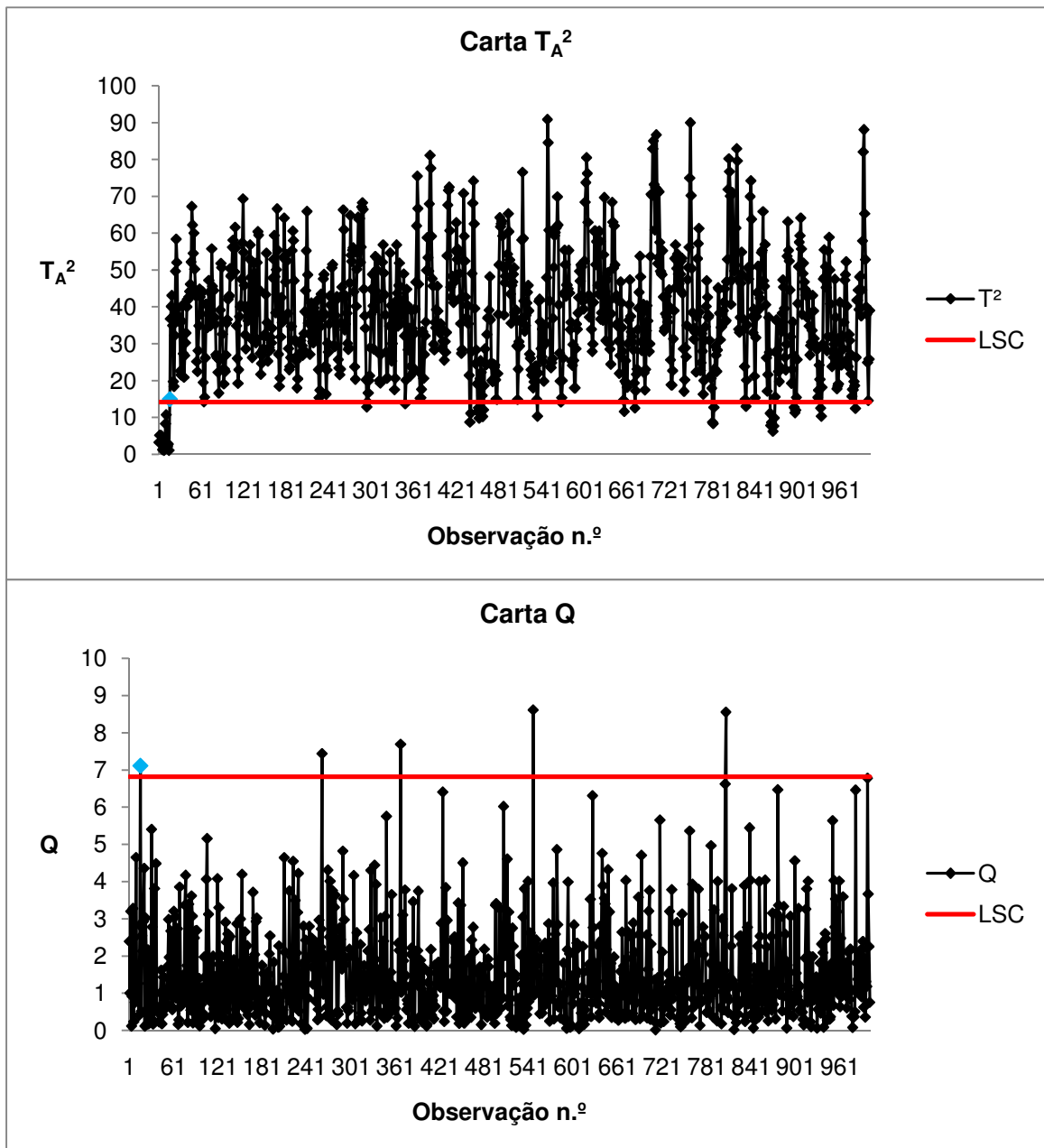


Figura 4.10 – Carta  $T^2$





**Figura 4.11 – Carta DPCA**

Analisando as Figuras 4.10 e 4.11, verifica-se que a tanto a carta  $T_A^2$  como a carta  $Q$  detectam a existência de um falso alarme, no instante anterior à introdução da perturbação, identificado a azul. Já a carta  $T^2$ , não detecta qualquer falso alarme.

### 4.3. Metodologias de Identificação de Causas Especiais em Cartas Multivariadas

Numa carta de controlo multivariada, nem sempre é fácil determinar a causa de uma situação de fora de controlo, pois esta tanto pode estar relacionada com uma única variável como a um subconjunto de variáveis. Assim, neste ponto pretende-se estabelecer uma comparação entre três métodos desenvolvidos para interpretação de causas especiais, quando o processo é sujeito a um tipo de perturbação. Este estudo será realizado com base no esquema apresentado no Capítulo 3 (Figura 3.6), ou seja, a perturbação será introduzida numa das variáveis, para posterior realização das duas cartas de controlo e aplicação dos diferentes métodos que permitirão identificar a/as variável/eis responsável/eis pela causa especial de variação gerada pela perturbação introduzida. Para a carta  $T^2$  será utilizado o método da estatística  $d$  e o método de decomposição em componentes ortogonais e para a carta DPCA será usado o gráfico de contribuições.

Relativamente à introdução da perturbação, esta vai ser feita pela adição de um valor de  $\delta\sigma_\varepsilon$  a cada um dos valores gerados da série de ruído branco ( $\varepsilon_t$ ), num dado instante. Por outro lado, como se pretende realizar uma comparação entre as duas cartas, a perturbação deve ser introduzida no mesmo instante, para as duas cartas, um pouco afastado do início, com uma magnitude de perturbação mínima ( $\delta = 1$  ou  $\delta = 1,5$ ) e que possa ser detectada tanto na carta  $T^2$  como na carta DPCA. Já em relação ao tipo de perturbação, esta será igual ao estudo anterior, ou seja, perturbação tipo degrau. Assim, os instantes a seguir à introdução da perturbação também serão alterados, podendo-se detectar a perturbação, nas duas cartas, no próprio instante ou nos instantes seguintes.

No que diz respeito à construção das duas cartas, como se pretende introduzir uma perturbação e verificar como é que estas reagem, os parâmetros (matriz covariância, limites de controlo, valores próprios e vectores próprios) que permitem construir estas cartas serão fixos, isto é, determinados sem a presença da perturbação. Por outro lado, pretende-se manter a mesma condição que no estudo anterior, isto é, que as duas cartas possuam o mesmo valor de  $ARL_{EC}$  igual a 370. Uma vez determinados os parâmetros, introduz-se a perturbação numa das variáveis e constróem-se as cartas de controlo.

A análise das cartas de controlo, só será feita a partir do instante que se introduziu a perturbação. Assim, caso haja pontos, antes deste instante, fora dos limites de controlo, estes não serão analisados por se tratar de um falso alarme. Por outro lado, deve-se escolher o ponto fora dos limites mais perto do instante em que se introduziu a perturbação pois caso se escolha um ponto mais afastado, este pode ter saído por um acumular de factores e não pela perturbação imposta. Deste modo, devem-se realizar várias simulações e escolher aquela que se considerar mais adequada, ou seja, que apresentar um ponto fora devido à perturbação, mais perto do instante em que esta foi introduzida.

Todas as simulações foram efectuadas com a perturbação introduzida no mesmo instante e na mesma variável, alterando-se apenas a magnitude da perturbação ( $\delta$ ). A escolha da melhor simulação foi difícil pois nenhuma ia ao encontro das tabelas do ARL e SDRL com alterações na média do processo (Tabelas 4.4 e 4.6), tendo-se escolhido a simulação que apresentou, nas duas cartas, o ponto fora mais perto do instante em que se introduziu a perturbação.

Na simulação escolhida, as séries de ruído branco para cada variável foram geradas para 5000 observações e a perturbação foi introduzida no instante  $t = 30$ . De seguida apresentam-se as duas cartas de controlo ( $T^2$  e DPCA), após a introdução da perturbação e os resultados obtidos aquando da aplicação dos três métodos (estatística  $d$ , decomposição em componentes ortogonais e gráfico de contribuição).

#### 4.3.1 Construção da Carta $T^2$ e Aplicação do Método da Estatística $d$ e do Método de Decomposição em Componentes Ortogonais

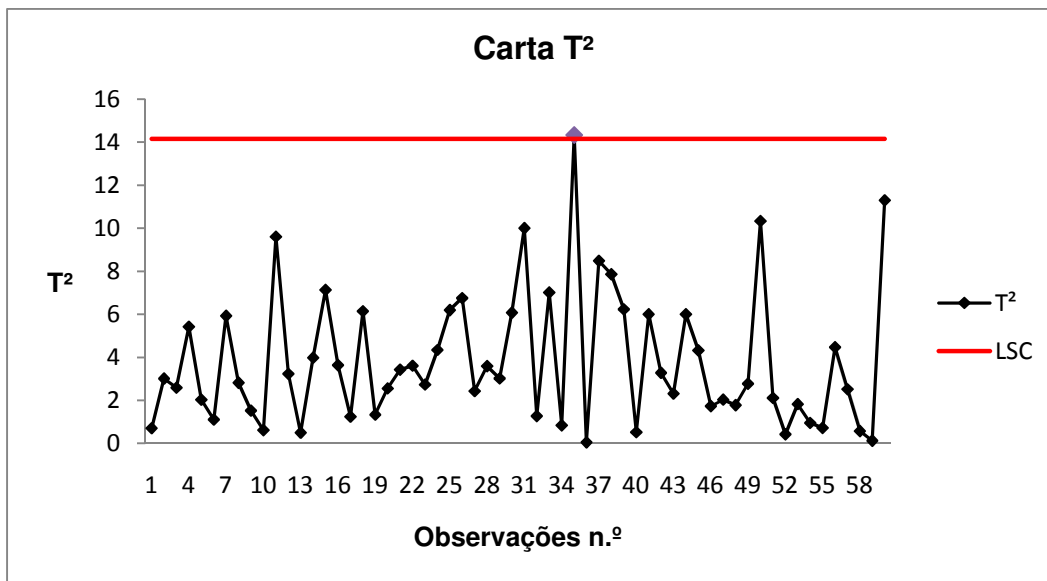
A construção da carta  $T^2$  foi feita, tendo em conta as seguintes condições:

- ✓ A matriz de covariância ( $S$ ) foi determinada, sem a presença da perturbação, com recurso ao *software* Excel;
- ✓ O limite para a carta  $T^2$  (expressão 2.31) foi determinado para um nível de significância ( $\alpha$ ) de 0,27%, de forma a obter-se um  $ARL_{EC} = 370$ .

Apresenta-se na Tabela 4.10 os valores de  $s_{ij}$  da matriz covariância ( $S$ ). A carta  $T^2$  encontra-se representada no ANEXO I (Figura I.1). Pela observação da carta  $T^2$ , representada na Figura I.1, verifica-se a existência de vários pontos acima do limite superior de controlo ( $LSC_{T^2} = 14,156$ ). Escolheu-se para análise o ponto fora dos limites mais perto do instante em que se introduziu a perturbação, sendo neste caso o ponto nº 35 ( $T^2_{35} = 14,335$ ) como se pode verificar na Figura 4.12, onde estão representados os primeiros 60 pontos da carta  $T^2$ .

**Tabela 4.10** – Valores de  $s_{ij}$  da matriz de covariância ( $S$ )

$s_{11}$	$s_{22}$	$s_{33}$	$s_{12}$	$s_{13}$	$s_{23}$
67,5657	9,4564	110,6610	0,3337	1,6338	0,6574



**Figura 4.12** – Carta T<sup>2</sup> para os primeiros 60 pontos

Para se identificar a/as variável/eis responsável/eis pela situação de fora de controlo, devido à perturbação imposta, utilizou-se o método da estatística  $d$  e o método de decomposição em componentes ortogonais. Todos os cálculos foram efectuados com recurso ao *software* Excel.

### **Estatística $d$**

Este método foi realizado com base nas expressões 2.42 a 2.43, tendo em conta os valores da matriz covariância (Tabela 4.10), os valores de cada variável correspondentes ao instante 35 e a média de cada variável.

Apresenta-se na Tabela 4.11 o resumo do estudo realizado para o ponto n.º 35, de forma a determinar que variável/eis é/são responsável/eis pela causa especial de variação.

**Tabela 4.11** – Estatística  $d$ 

<b>n.º</b>	<b><math>T^2</math></b>	<b><math>T^2_{(i)}</math></b>	<b><math>d_i = T^2 - T^2_{(i)}</math></b>	<b>Valor crítico (<math>\chi^2_{0,0027;1}</math>)</b>
35	14,335	$T^2_{(1)} = 3,612$	$d_1 = 10,723$	8,999
		$T^2_{(2)} = 10,891$	$d_2 = 3,444$	8,999
		$T^2_{(3)} = 10,889$	$d_3 = 3,446$	8,999

Analisando a Tabela 4.11, conclui-se que a variável 1 é responsável pela causa especial de variação ( $d_1 >$  valor crítico), indicando assim que houve uma alteração na média desta variável.

#### **Decomposição em componentes ortogonais**

Este método foi realizado com base nas expressões 2.44 a 2.56, tendo em conta os valores da matriz covariância (Tabela 4.10), os valores de cada variável correspondentes ao instante 35 e a média de cada variável.

Na Tabela 4.12 apresenta-se o resumo do estudo efectuado para o ponto nº 35, de forma a determinar que variável/eis e/ou correlação entre as variáveis é responsável/eis pela causa especial de variação. Nesta Tabela, “nd” significa “não determinado”.

**Tabela 4.12** – Decomposição em componentes ortogonais

n.º	$T^2$	Termos incondicionais ( $T_j^2$ )	$(T^2 - T_j^2)$	Termos condicionais	Valor crítico [(m + 1)/m · $F_{\alpha;1;m-1}$ ]
35	14,335	$T_1^2 = 10,887$	3,448	$T_{2,1}^2 = \text{nd}$	9,009
				$T_{3,1,2}^2 = \text{nd}$	9,009
				$T_{3,1}^2 = \text{nd}$	9,009
				$T_{2,1,3}^2 = \text{nd}$	9,009
		$T_2^2 = 3,610$	10,725	$T_{1,2}^2 = 10,724$	9,009
				$T_{3,2,1}^2 = 5,23 \times 10^{-6}$	9,009
				$T_{3,2}^2 = 0,004$	9,009
				$T_{1,2,3}^2 = 10,720$	9,009
		$T_3^2 = 0,010$	14,325	$T_{1,3}^2 = 10,904$	9,009
				$T_{2,3,1}^2 = 0,002$	9,009
				$T_{2,3}^2 = 3,604$	9,009
				$T_{1,3,2}^2 = 10,720$	9,009

Analisando os resultados obtidos na Tabela 4.12, conclui-se que a causa especial de variação é devido a:

- ✓ Variável 1;
- ✓ Correlação entre a variável 1 e a variável 2;
- ✓ Correlação entre a variável 1 e a variável 3;
- ✓ Correlação entre a variável 1 e as variáveis 2 e 3.

Como se pode verificar, este método é muito mais abrangente, permitindo assim concluir que afinal a causa especial de variação não foi só devido a uma alteração na variável 1 (variável

que se introduziu a perturbação) como também foi devido a uma alteração da correlação entre a variável 1 e as restantes variáveis.

Comparando os dois métodos, constata-se que o método de decomposição em componentes ortogonais, apesar da sua aplicação ser mais complexa, é mais completo que o método da estatística  $d$  na medida que não só permite verificar se a situação de fora de controlo é devido a alteração de uma ou mais variáveis consideradas isoladamente como também se é devido a uma alteração da correlação entre variáveis.

### 4.3.2 Construção da Carta DPCA e Aplicação do Gráfico de Contribuição

Para a construção das cartas  $T_A^2$  e  $Q$ , teve-se por base as seguintes condições:

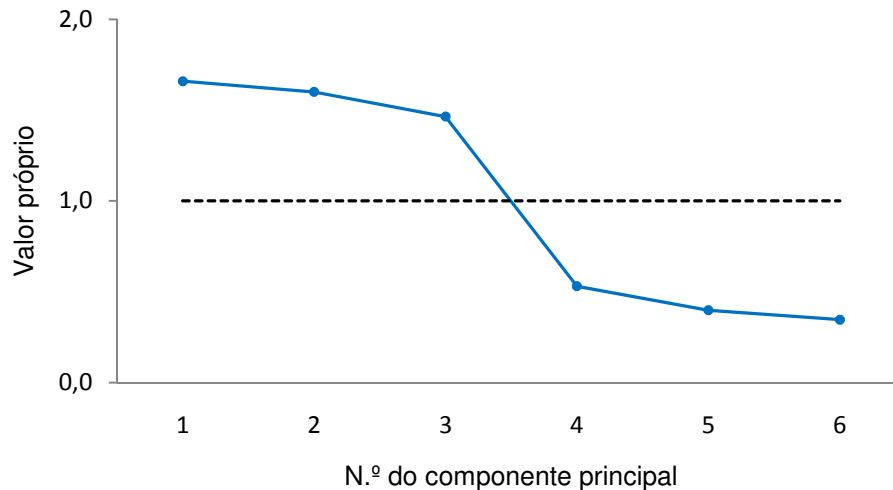
- ✓ A normalização dos dados para que todas as variáveis apresentassem média igual a zero e variância unitária feita no *software* Excel, através da fórmula:

$$X_{norm} = (X - \bar{X}) / DP_{original} ;$$

- ✓ A matriz de dados foi construída com um atraso de uma unidade de tempo, obtendo-se deste modo uma matriz com 6 colunas e 5000 linhas. A ordem pela qual as observações das variáveis normalizadas foram dispostas na matriz de dados corresponde à ordem da matriz dos vectores próprios;
- ✓ Os vectores próprios e valores próprios foram determinados sem a presença da perturbação, através do *software* STATISTICA;
- ✓ O número de componentes a considerar foi determinado através do critério de *Kaiser* que estabelece que apenas devem ser retidas as componentes principais com valores próprios superiores a um;
- ✓ O limite para a carta  $T_A^2$  (expressão 2.31) foi determinado para um nível de significância ( $\alpha$ ) de 0,27%, de forma a obter-se um  $ARL_{EC} = 370$ .
- ✓ O limite para a carta  $Q$  (expressão 2.69) foi determinado para um nível de significância ( $\alpha$ ) de 0,16%, de forma a obter um  $ARL_{EC} = 370$  e mantêve-se fixo.

Na Figura 4.13 apresenta-se, graficamente, a aplicação do critério de *Kaiser* de forma a indicar o número de componentes principais que se considerou neste modelo. Os valores próprios

correspondentes aos três primeiros componentes principais, a percentagem de variância retida por cada um e a percentagem de variância cumulativa encontram-se na Tabela 4.13. Analisando a Tabela 4.13, verifica-se que os três primeiros componentes principais explicam 79 % da variabilidade dos dados, o que é considerado satisfatório.



**Figura 4.13** – Critério de *Kaiser* para identificação do número de componentes principais

**Tabela 4.13** – Valores próprios correspondentes aos três componentes principais retidos e percentagem de variância cumulativa dos mesmos

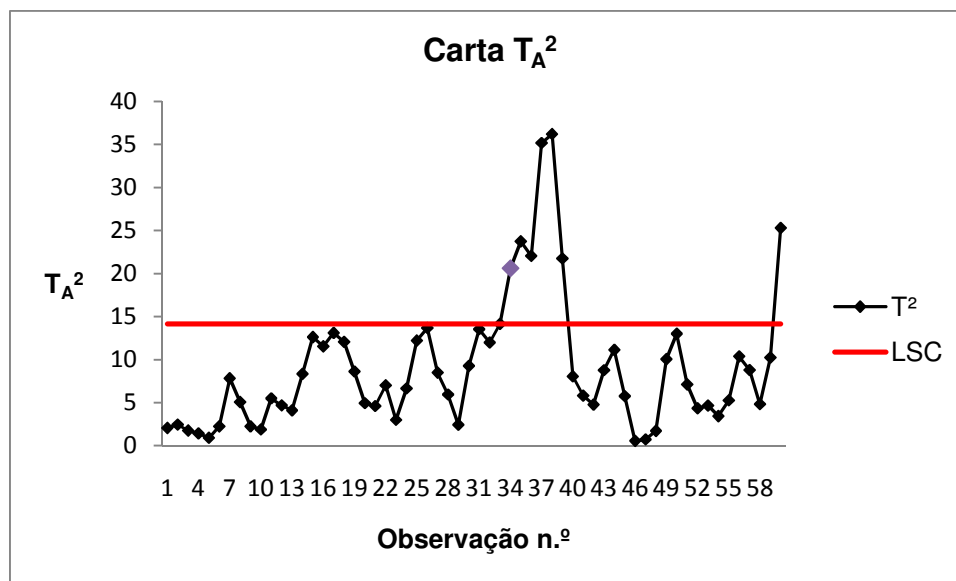
Componentes principais	Valores próprios	% de variância	% de variância cumulativa
1	1,659	27,649	27,649
2	1,600	26,661	54,310
3	1,464	24,401	78,710

A matriz dos vectores próprios para os três componentes principais retidos no modelo, está apresentada na Tabela 4.14. Já a carta DPCA encontra-se representada no ANEXO I (Figura I.2). Pela análise das cartas  $T_A^2$  e  $Q$ , representada na Figura I.2 (ANEXO I), verifica-se, a existência de vários pontos acima do limite superior de controlo, o que significa que houve uma alteração na estrutura dos componentes principais dinâmicos. Escolheu-se para análise o ponto fora dos limites mais perto do instante em que se introduziu a perturbação, sendo neste caso o ponto nº 34 ( $T_{34}^2 = 20,603$ ) referente a carta  $T_A^2$  (Figura 4.14).



**Tabela 4.14** – Matriz dos vectores próprios ( $P$ )

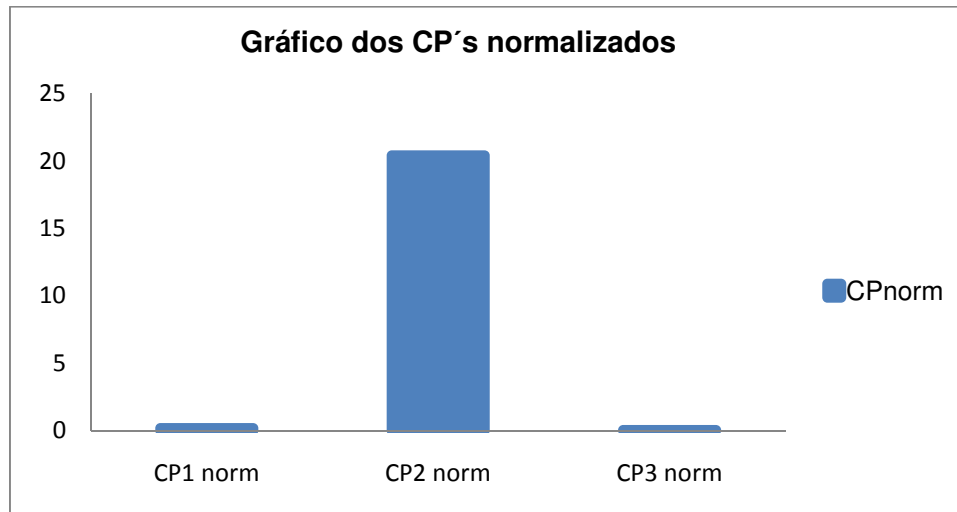
	Componentes n.º		
	1	2	3
$X1_t$	-0,166571	-0,681003	0,092473
$X2_t$	0,680034	-0,148945	0,125097
$X3_t$	-0,100727	0,118550	0,689433
$X1_{t-1}$	-0,173675	-0,681185	0,075804
$X2_{t-1}$	0,681198	-0,163540	0,094660
$X3_{t-1}$	-0,073948	0,096118	0,696976

**Figura 4.14** – Carta  $T_A^2$  para os primeiros 60 pontos

Para se identificar a/as variável/eis responsável/eis pela situação de fora de controlo, devido à perturbação introduzida, utilizou-se os gráficos de contribuição. Todos os cálculos foram efectuados com recurso ao *software* Excel.

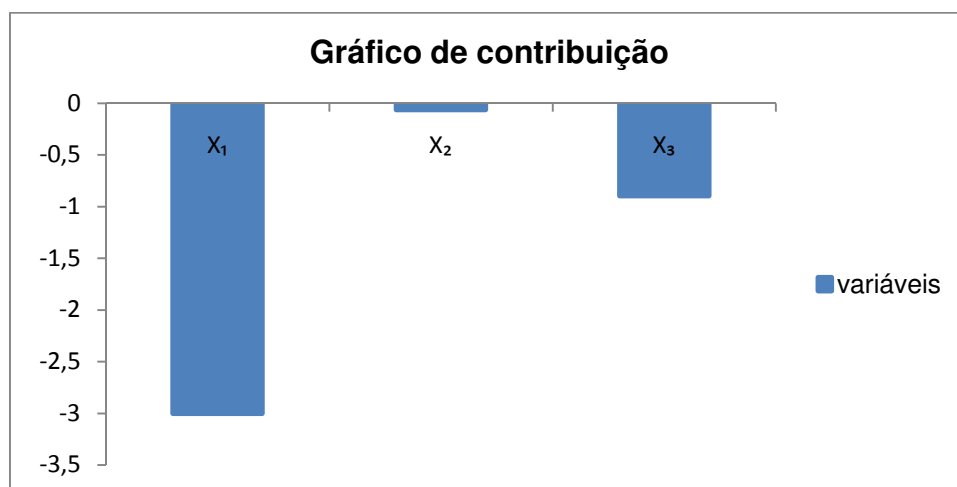
Como a situação de fora de controlo (ponto n.º 34 acima do limite superior) foi detectada na carta  $T_A^2$ , primeiro contruiu-se um gráfico de barras (Figura 4.15) dos componentes principais

normalizados de forma a identificar qual o componente principal responsável pela causa especial. Os componentes principais normalizados foram determinados pela expressão  $(t_i/\lambda_i)$ .



**Figura 4.15** – Gráfico dos componentes principais normalizados

Analisando o gráfico da Figura 4.15, constata-se que o segundo componente principal é responsável pela situação de fora de controlo pois apresenta o maior valor normalizado. Uma vez identificado o componente principal responsável pela situação de fora de controlo, determinou-se a contribuição das variáveis para este componente principal, através da expressão 2.71. Com estes valores construiu-se o gráfico de contribuição (Figura 4.16) de forma a indicar como é que cada variável envolvida no cálculo deste componente principal contribui para o sinal fora de controlo.



**Figura 4.16** – Gráfico de contribuição das variáveis

Observando a Figura 4.16, verifica-se que a variável 1 é a mais envolvida na composição do componente principal responsável pela causa especial pois é a que apresenta um valor de contribuição mais elevado. Assim, e uma vez que a variável 1 apresenta o mesmo sinal que o componente principal, esta deve ser investigada por contribuir para a situação de fora de controlo.

Como se pode verificar, através deste método facilmente se identifica que foi na variável 1 que ocorreu a perturbação, uma vez que esta é a responsável pela situação de fora de controlo indicada.

#### 4.4. Vantagens e Desvantagens das Duas Cartas de Controlo

Face ao que foi até então exposto, pretende-se, neste ponto, dar uma visão prática da aplicação destas cartas em ambiente industrial, através da apresentação das principais vantagens e desvantagens identificadas para cada uma das cartas de controlo em estudo.

##### Vantagens da carta $T^2$

- ✓ Permite a monitorização simultânea de duas ou mais características da qualidade relacionadas;
- ✓ Para estabelecer os limites de controlo, considerando um  $ARL_{EC}$  específico, não é necessário recorrer a modelos de simulação;
- ✓ Facilidade de aplicação;
- ✓ Apresenta um desempenho razoável quando se pretende controlar um número pequeno de variáveis;
- ✓ Pode ser aplicada a dado autocorrelacionados;
- ✓ É necessário interpretar apenas uma carta de controlo.

**Desvantagens da carta  $T^2$** 

- ✓ Tende a perder a eficácia no que respeita à detecção de alterações no processo, quando o número de variáveis a controlar é elevado;
- ✓ Não é muito eficiente na detecção de alterações pequenas a moderadas nos parâmetros do processo, uma vez que só usa a informação contida na amostra, mais recentemente inspecionada;
- ✓ A sua utilização pressupõe variáveis independentes entre si e Normalmente distribuídas, sendo necessário proceder-se a uma transformação de dados ou a modelação do processo (dados autocorrelacionados) quando esta suposição é violada, o que é frequente em ambiente industrial, especialmente na indústria química.

**Vantagens da carta DPCA**

- ✓ Permite o controlo simultâneo de um grande número de variáveis;
- ✓ Simplifica a análise dos dados, uma vez que permite reduzir o número de variáveis a controlar, sem perda significativa de informação;
- ✓ Fornece as ferramentas adequadas para identificar as variáveis mais importantes no espaço das componentes principais;
- ✓ É muito sensível na detecção de alterações na média do processo;
- ✓ Caso os dados exibam autocorrelação, esta carta permite eliminar a autocorrelação sem ter-se de modelar o processo, uma vez que a sua utilização não pressupõe variáveis Normalmente distribuídas;
- ✓ É uma ferramenta eficiente na monitorização de processos muito complexos, com grande número de variáveis.

### **Desvantagens da carta DPCA**

- ✓ Embora esteja disponível no mercado várias aplicações informáticas que permitam efectuar análise de componentes principais, a construção da carta exige um conhecimento específico na área da estatística, não sendo, por isso, de fácil aplicação;
- ✓ Utiliza duas cartas que têm de ser interpretadas em conjunto;
- ✓ A determinação dos limites de controlo para um  $ARL_{EC}$  específico exige o recurso à construção de modelos de simulação.

## **4.5. Síntese**

Neste capítulo realizaram-se dois estudos diferentes, de forma a indicar qual a carta multivariada com dados autocorrelacionados que melhor desempenho apresenta e quais os métodos a utilizar caso seja detectado um sinal de fora de controlo, associado a uma causa especial de variação. Estes dois estudos tiveram por base os fundamentos teóricos apresentados no capítulo 2 bem como os esquemas apresentados no capítulo 3.

O primeiro estudo teve, como principal objectivo, estabelecer uma comparação entre o desempenho da carta  $T^2$  e da carta DPCA, quando o processo é sujeito à introdução de uma perturbação no parâmetro médio do processo, do tipo degrau. Já o segundo estudo, teve como principal objectivo estabelecer uma comparação entre os três métodos desenvolvidos para interpretação de causas especiais, quando o processo é sujeito a um tipo de perturbação.

As constatações que se apresentam de seguida dizem respeito aos dois estudos realizados:

- ✓ Da análise do desempenho das duas cartas, conclui-se, com base nos valores de ARL obtidos, que a carta DPCA é a mais eficaz pois apresenta o melhor valor de ARL, permitindo assim detectar com maior rapidez alterações reduzidas, moderadas ou elevadas na média do processo. Por outro lado, esta carta também é a que apresenta um valor mais baixo de SDRL;
- ✓ A carta DPCA, após a introdução da perturbação, mostrou-se muito mais sensível para detectar as alterações na média do processo do que a carta  $T^2$ ;
- ✓ Os três métodos (estatística  $d$ , decomposição em componentes ortogonais e gráficos de contribuição) mostraram-se eficazes na identificação da variável em que ocorreu a perturbação;

- ✓ Em termos de facilidade de aplicação, o método da estatística  $d$  e o gráfico de contribuição foram os mais acessíveis;
- ✓ A análise de componentes principais juntamente com o gráfico de contribuição mostrou-se de grande utilidade na identificação da causa especial no processo produtivo;
- ✓ Dos dois métodos aplicados à carta  $T^2$ , conclui-se que o método de decomposição em componentes ortogonais é o mais abrangente pois permite verificar se a causa especial de variação é devido a uma alteração na média do processo como também se é devido a uma alteração da correlação entre variáveis.

Por outro lado, com a realização dos dois estudos conseguiu-se identificar-se quais as principais vantagens e desvantagens das duas cartas multivariadas (carta  $T^2$  e carta DPCA) na óptica do utilizador que vai aplicar a carta de controlo em ambiente industrial.

## 5. Conclusão e Sugestões para Trabalhos Futuros

Neste capítulo, pretende-se apresentar as principais conclusões que foram retiradas de cada um dos capítulos que compõem este trabalho de investigação. Por fim, serão apresentadas sugestões para possíveis trabalhos futuros, dentro da temática do controlo estatístico multivariado do processo.

### 5.1. Conclusões gerais

Com este trabalho de investigação, pretendeu-se realizar uma análise comparativa do desempenho de duas cartas de controlo multivariadas, com dados autocorrelacionados, de forma a indicar qual delas é a mais adequada. As cartas de controlo escolhidas para o desenvolvimento deste estudo foram a carta  $T^2$ , por ser a mais conhecida e puder ser aplicada a dados autocorrelacionados, e a carta baseada na análise de componentes principais dinâmicas (DPCA), por permitir reduzir o número de variáveis consideradas e também puder ser aplicada a dados autocorrelacionados.

O desenvolvimento deste estudo foi feito segundo duas perspectivas, a teórica e a prática. A parte teórica, onde constam os fundamentos teóricos necessários ao desenvolvimento da metodologia proposta, foi realizada com recurso a vários artigos e livros científicos referentes às várias temáticas ligadas à qualidade. A parte prática procurou, de certa forma, dar validade os desenvolvimentos propostos.

O capítulo dedicado ao estudo das cartas de controlo teve, como objectivo, identificar a melhor carta de controlo e os métodos que devem ser utilizados na detecção de causas especiais. O primeiro estudo efectuado teve, como principal objectivo, estabelecer uma comparação entre o desempenho das duas cartas de controlo, quando o processo é sujeito a um tipo de perturbação numa ou mais variáveis em simultâneo. Estas cartas foram construídas tendo por base o mesmo modelo matemático, AR(1) e as mesmas variáveis. Este estudo foi realizado através de simulação, no qual se concluiu que a carta baseada na análise de componentes principais dinâmicas (DPCA) é melhor que a carta  $T^2$  em termos de desempenho de ARL, ou seja, esta carta é muito mais rápida na detecção de alterações reduzidas, moderadas ou elevadas no parâmetro da média do processo e também é a que apresenta um valor de SDRL mais baixo.

O segundo estudo, teve como principal objectivo estabelecer uma comparação entre os três métodos desenvolvidos para interpretação de causas especiais, quando o processo é sujeito a um tipo de perturbação. Pela realização dos três métodos, concluiu-se que o método de

decomposição em componentes ortogonais (aplicado a carta  $T^2$ ) e os gráficos de contribuição (aplicado a carta DPCA) são os mais apropriados quando se pretende identificar as variáveis responsáveis por uma situação de fora de controlo. O método de decomposição em componentes ortogonais revelou-se numa ferramenta melhor que o método da estatística  $d$  uma vez que permite verificar se a causa especial de variação é devido a alteração de uma variável e/ou a alteração da correlação entre variáveis.

Para além destas conclusões, também se retiraram algumas conclusões globais referentes à investigação científica realizada neste trabalho:

- ✓ A verificação da existência de autocorrelação antes de iniciar o controlo estatístico do processo é de extrema importância uma vez que permite evitar o uso de cartas de controlo incorrectas, pois como se pode constatar, a autocorrelação tem um impacto acentuado no desempenho das cartas de controlo;
- ✓ O desenvolvimento de metodologias, recorrendo a fluxogramas, é muito útil para realização do que se pretende aplicar na prática;
- ✓ Destacam-se os aspectos positivos e negativos da utilização de cada uma das cartas através de uma aplicação real;
- ✓ As conclusões retiradas do estudo do ARL, para a carta DPCA, foram uma mais valia uma vez que na literatura encontram-se poucos estudos desta natureza;
- ✓ A carta DPCA demonstrou-se numa excelente ferramenta estatística para monitorizar a qualidade de um processo, quando se pretende detectar alterações no parâmetro médio do processo;
- ✓ Uma vez que a carta DPCA apresenta um melhor desempenho que a carta  $T^2$ , recomenda-se o uso desta carta como alternativa à carta  $T^2$ ;
- ✓ A construção da carta DPCA pode ser mais complicada que a da carta  $T^2$ , no entanto, a melhoria substancial na redução do ARL deve justificar a sua aplicação;
- ✓ A interpretação de causas especiais, que nem sempre é fácil em cartas multivariadas, pode ser devidamente tratada através da aplicação do método de decomposição em componentes ortogonais (carta  $T^2$ ) e dos gráficos de contribuição (carta DPCA);



- ✓ Embora já exista no mercado *software* que permite efectuar a análise de componentes principais e, também a aplicação de cartas multivariadas baseadas em PCA, seria possível adaptar estes programas a uma carta DPCA.

Como se pode verificar, este trabalho apresenta resultados satisfatórios, alcançando-se assim todos os objectivos pretendidos.

## 5.2. Sugestões para Trabalhos Futuros

Como qualquer trabalho de investigação científica, várias sugestões vão surgindo mas que não foram possíveis aprofundar com o presente trabalho. Assim, em continuidade ao desenvolvimento deste estudo, apresentam-se várias sugestões para trabalhos futuros.

Relativamente aos modelos matemáticos que foram utilizados, considerou-se que seria interessante usar outros modelos e tentar determinar um padrão entre a estrutura autocorrelacionada dos dados e a composição da matriz dinâmica. Por exemplo utilizar um modelo ARMA (1,1) e um modelo AR(1) e verificar qual deles seria o mais adequado em termos de valores de ARL.

Uma vez que as observações não se comportam de forma independente, seria apropriado analisar o comportamento das cartas quando ocorrem alterações na dinâmica do processo, isto é, alterações nos parâmetros do modelo matemático ARIMA.

Uma outra área de interesse científico seria aprofundar as cartas CUSUM e EWMA multivariadas, para processos com dados autocorrelacionados, e compará-las com a carta  $T^2$  e a carta baseada em análise de componentes principais (DPCA), através do valor de ARL.

Seria importante realizar mais estudos no sentido de perceber de que forma será possível eliminar a autocorrelação dos dados, tendo sempre em vista a componente prática.

## Referências Bibliográficas

- Alwan, L. C., e Roberts, H. V. (1988). Time-Series Modeling for Statistical Process Control. *Journal of Business & Economic Statistics* , 6 (1), pp. 87-95.
- Aparisi, F., Champ, C. W., e Garcia-Diaz, J. C. (2004). A Performance Analysis of Hotelling  $X^2$  Control Chart with Supplementary Run Rules. *Quality Eng.* , 16 (3), pp. 359-368.
- Bersimis, S., Psarakis, S., e Panaretos, J. (2007). Multivariate Statistical Process Control Charts: An Overview. *Quality and Reliability Engineering International* , 23, pp. 517-543.
- Box, G. E., Jenkins, G. M., e Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control* (4ª ed.). Wiley.
- Crosier, R. B. (1988). Multivariate Generalizations of Cumulative Sum Quality Control Schemes. *Technometrics* , 30 (3), pp. 291-303.
- English, J. R., Lee, C. S., Martin, T. W., e Tilmon, C. (2000). Detecting Changes in Autoregressive Processes with X-Bar and EWMA Charts. *IIE Transactions* , 32 (12), pp. 1103-1113.
- Ferrer, A. (2007). Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. *Quality Engineering* , 19, pp. 311-325.
- Ghute, V. B., e Shirke, D. T. (2008). A Multivariate Synthetic Control Chart for Monitoring Process Mean Vector. *Communications in Statistics -Theory and Methods* , 37, pp. 2136–2148.
- Harris, T. J., e Ross, W. H. (1991). Statistical Process Control Procedures for Correlated Observations. *The Canadian Journal of Chemical Engineering* , 69, pp. 48-57.
- Jackson, J. E. (1985). Multivariate Quality Control. *Communications in Statistics - Theory and Methods* , 14 (11), pp. 2657-2688.
- Javaheri, A., e Houshmand, A. (2001). Average Run Length Comparison of Multivariate Control Charts. *Journal of Statistical Computation and Simulation* , 69, pp. 125 - 140.
- Kano, M., Hasebe, S., Hashimoto, I., e Ohno, H. (2001). A New Multivariate Statistical Process Monitoring Method Using Principal Component Analysis. *Computers and Chemical Engineering*, 25, pp. 1103-1113.

- Kourti, T. (2005). Application of Latent Variable Methods to Process Control and Multivariate Statistical Process Control in Industry. *International Journal of Adaptive Control and Signal Processing*.
- Kourti, T., e MacGregor, F. (1995). Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods. *Chemometrics and Intelligent Laboratory Systems*, 28, pp. 3-21.
- Ku, W., Storer, R. H., e Georgakis, C. (1995). Disturbance Detection and Isolation by Dynamic Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 30, pp. 179-196.
- Lin, W. S., e Adams, B. M. (1996). Combined Control Chart for Forecast-Based Monitoring Schemes. *Journal of Quality Technology*, 28 (3), pp. 289-301.
- Ljung, L. (2002). *System Identification Toolbox: For Use with MatLab*.
- Lowry, C. A., e Montgomery, D. C. (1995). A Review of Multivariate Control Charts. *IIE Transactions*, 27, pp. 800-810.
- Lowry, C. A., Woodall, W. H., Champ, C. W., e Rigdon, S. E. (1992). Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, 34, pp. 46-53.
- Lu, C. W., e Reynolds, M. R. (1999a). Control Chart for Monitoring the Mean and Variance of Autocorrelated Processes. *Journal of Quality Technology*, 31 (3), pp. 259-274.
- Lu, C. W., e Reynolds, M. R. (2001). CUSUM Chart for Monitoring an Autocorrelated Process. *Journal of Quality Technology*, 33 (3), pp. 316-334.
- Lu, C. W., e Reynolds, M. R. (1999b). EWMA Control Chart for Monitoring the Mean of Autocorrelated Processes. *Journal of Quality Technology*, 31 (2), pp. 166-188.
- MacGregor, J., e Kourti, T. (1995). Statistical Process Control of Multivariate Processes. *Control Engineering Practice*, 3, pp. 403-414.
- Mason, R. L., Tracy, N. D., e Young, J. C. (1997). A Practical Approach for Interpreting Multivariate  $T^2$  Control Chart Signals. *Journal of Quality Technology*, 29, pp. 396-406.
- Mason, R. L., Tracy, N. D., e Young, J. C. (1995). Decomposition of  $T^2$  for Multivariate Control Chart Interpretation. *Journal of Quality Technology*, 27, pp. 99-108.
- Matos, A. S. (2006). Engenharia de Controlo do Processo e Controlo Estatístico da Qualidade: Metodologia de Integração Aplicada na Indústria da Pasta de Papel. *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*.

- Miranda, L. M. (2002). Qualidade do Processo Produtivo da Pasta de Papel. *Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa*.
- Montgomery, D. C. (2005). *Introduction to Statistical Quality Control* (5ª ed.). Jonh Wiley & Sons.
- Montgomery, D. C., e Mastrangelo, C. M. (1991). Some Statistical Process Control Methods of Autocorrelated Data. *Journal of Quality Technology*, 23 (3), pp. 179-193.
- Morais, V., e Vieira, C. (2006). *MATLAB 6 & 7 - Curso Completo*. Portugal.
- Murteira, B., Muller, D., & Turkman, K. (2000). *Análise de Sucessões Cronológicas*. MC.Granw-Hill. Portugal.
- Nomikos, P., e MacGregor, J. F. (1995). Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, 37 (1), pp. 41-59.
- Ostyn, B., Darius, P., Baerdemaeker, J., e Ketelaere, B. (2007). Statistical Monitoring of a Sealing Process by Means of Multivariate Accelerometer Data. *Quality Engineering*, 19, pp. 299-310.
- Pereira, Z. L., e Requeijo, J. G. (2008). *Qualidade: Planeamento e Controlo Estatístico de Processos*. Caparica: FCT- Fundação da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.
- Reynolds, M. R., e Lu, C. W. (1997). Control Chart for Monitoring Processes With Autocorrelated Data. *NonLinear Analysis, Theory, Methods Applications*, 30 (7), pp. 4059-4067.
- Runger, G. C., e Willemain, T. R. (1995). Model-Based and Model-Free Control of Autocorrelated Processes. *Journal of Quality Technology*, 27 (4), pp. 283-292.
- Runger, G. C., Alt, F. B., & Montgoremy, D. C. (1996). Contributors to a Multivariate Statistical Process Control Signal. *Communications in Statistics -Theory and Methods*, 25, pp. 2203-2213.
- Russell, E. L., Chiang, L. H., e Braatz, R. D. (2000). Fault Detection in Industrial Processes Using Canonical Variate Analysis and Dynamic Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 51, pp. 81-93.
- Tracy, N. D., Young, J. C., e Mason, R. L. (1992). Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, 24 (2), pp. 88-95.
- Tsung, F. (2000). Statistical Monitoring and Diagnosis of Automatic Controlled Processes Using Dynamic PCA. *International Journal of Production Research*, 38 (3), pp. 625-637.

- Vander Wiel, S. A. (1996a). Monitoring Processes That Wander Using Integrated Moving Average Models. *Technometrics* , 38 (2), pp. 139-151.
- Vasilopoulos, A. V., e Stamboulis, A. P. (1978). Modification of Control Chart Limits in the Presence of Data Correlation. *Journal of Quality Technology* , 10 (1), pp. 20-30.
- Wardell, D., Moskowitz, H., e Plante, R. (1994). Run-Length Distributions of Special-Cause Control Charts for Correlated Processes. *Technometrics* , 36 (1), pp. 3-27.
- Wieringer, J. E. (1999). Statistical Process Control for Serially Correlated Data. *Ph.D. Thesis University of Groningen* .
- Yang, J., e Hancock, W. M. (1990). Statistical Quality Control for Correlated Samples. *International Journal of Production Process* , 28 (3), pp. 595-608.
- Yashchin, E. (1993). Performance of CUSUM Control Schemes for Serially Correlated Observations. *Technometrics* , 35 (1), pp. 37-52.
- Zang, N. F. (1998). A Statistical Control Chart for Stationary Process Data. *Technometrics* , 40 (1), pp. 24-28.

## **ANEXOS**

### **ANEXO I – Aplicação de cartas multivariadas ao processo**



## ANEXO I – Aplicação de cartas de controlo multivariadas ao processo

### I.1. Valores de ARL e SDRL considerando alterações no parâmetro médio do processo

#### I.1.1. Estudo do desempenho da carta $T^2$

**Tabela I.1** – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo quando se introduz perturbações numa variável, considerando o estudo da carta  $T^2$

$\delta$	Carta $T^2$			
	Perturbação no $\varepsilon_2$		Perturbação no $\varepsilon_3$	
	ARL	SDRL	ARL	SDRL
0	369,85	369,59	370,5	368,59
0,5	232,74	231,92	232,69	231,67
1	91,13	90,06	91,92	91,4
1,5	35,16	34,42	35,54	35,71
2	15,39	14,79	15,22	14,76
2,5	7,56	7,02	7,63	7,09
3	4,33	3,81	4,32	3,79

**Tabela I.2** – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo quando se introduz perturbações em duas variáveis, considerando o estudo da carta  $T^2$

$\delta$	Carta $T^2$			
	Perturbação no $\varepsilon_1$ e $\varepsilon_3$		Perturbação no $\varepsilon_2$ e $\varepsilon_3$	
	ARL	SDRL	ARL	SDRL
0	370,03	367,15	370,49	368,39
0,5	160,98	160,28	162,23	161,25
1	41,27	40,75	41,56	40,89
1,5	12,78	12,26	12,74	12,21
2	5,14	4,62	5,16	4,62
2,5	2,69	2,13	2,69	2,14
3	1,73	1,12	1,75	1,15



### I.1.2. Estudo do desempenho da carta DPCA

**Tabela I.3** – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo quando se introduz uma perturbação numa variável, considerando o estudo da carta DPCA

$\delta$	Carta DPCA			
	Perturbação no $\varepsilon_2$		Perturbação no $\varepsilon_3$	
	ARL	SDRL	ARL	SDRL
0	369,42	364,73	370,23	370,16
0,5	85,73	82,89	112,33	111,03
1	17,33	14,74	24,44	22,88
1,5	6,36	4,34	7,63	6,43
2	3,43	1,89	3,53	2,56
2,5	2,29	1,07	2,08	1,27
3	1,68	0,71	1,44	0,71

**Tabela I.4** – Valores de ARL e SDRL, para diferentes alterações no parâmetro médio do processo quando se introduz uma perturbação em duas variáveis, considerando o estudo da carta DPCA

$\delta$	Carta DPCA			
	Perturbação no $\varepsilon_1$ e $\varepsilon_3$		Perturbação no $\varepsilon_2$ e $\varepsilon_3$	
	ARL	SDRL	ARL	SDRL
0	369	364,8	370,06	365,08
0,5	49,27	47,13	54,66	52,27
1	8,01	6,3	8,79	6,83
1,5	3,01	1,83	3,28	1,97
2	1,69	0,83	1,83	0,88
2,5	1,19	0,43	1,26	0,48
3	1,03	0,17	1,05	0,22

## I.2. Carta $T^2$

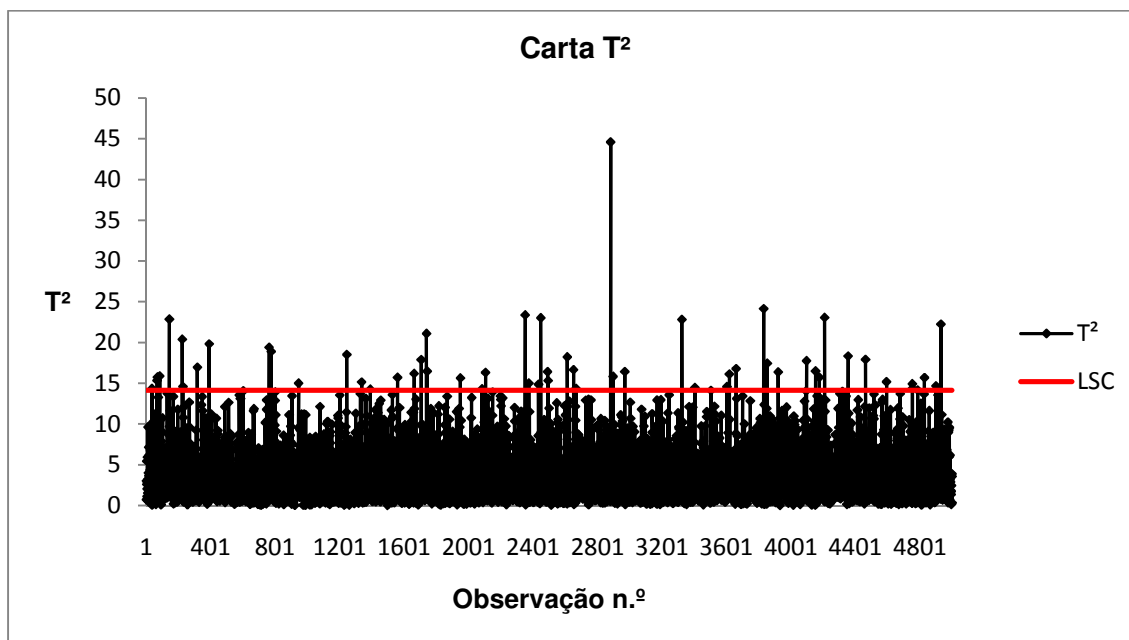
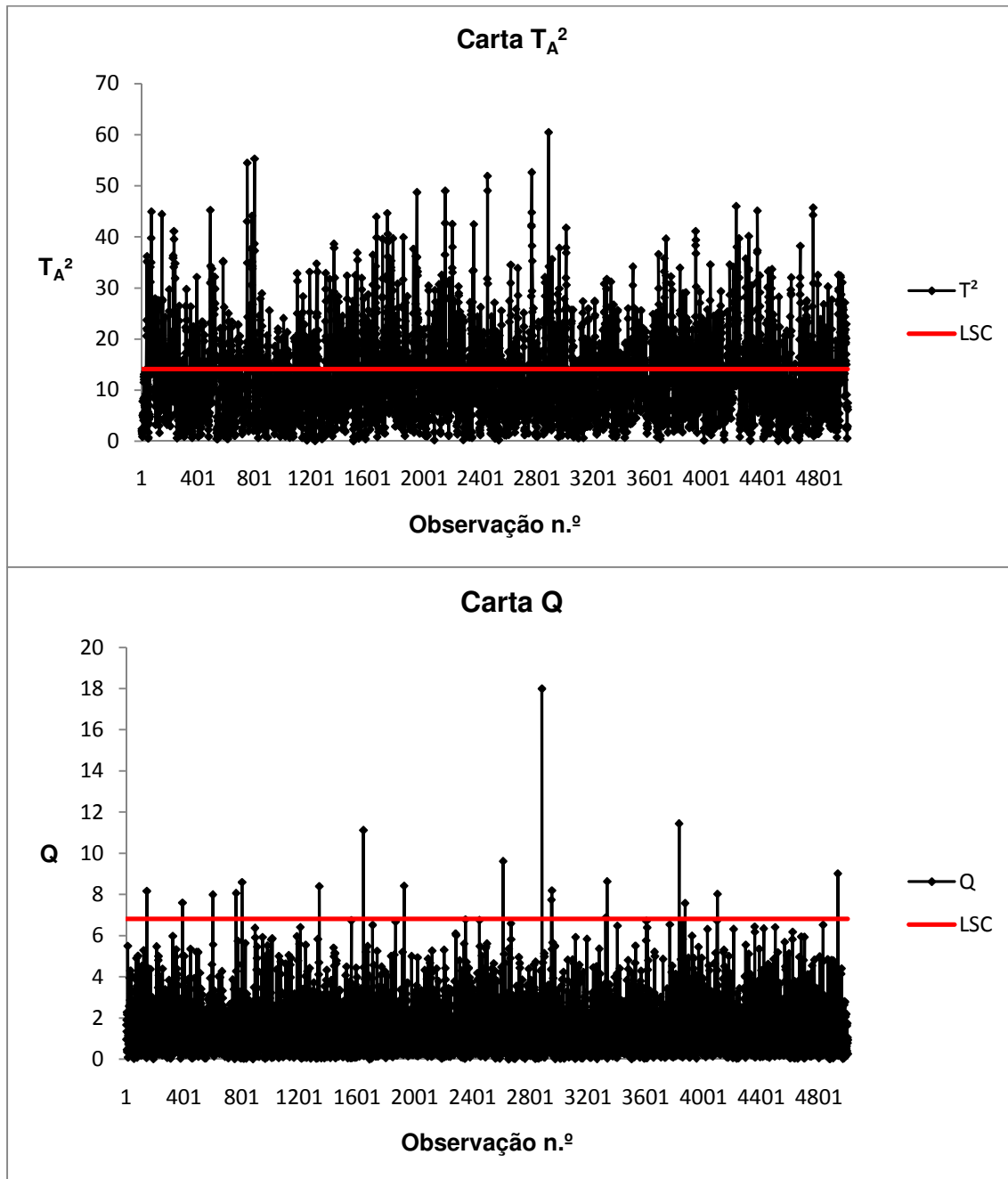


Figura I.1 – Carta  $T^2$

### I.3. Carta DPCA



**Figura I.2 – Carta DPCA**